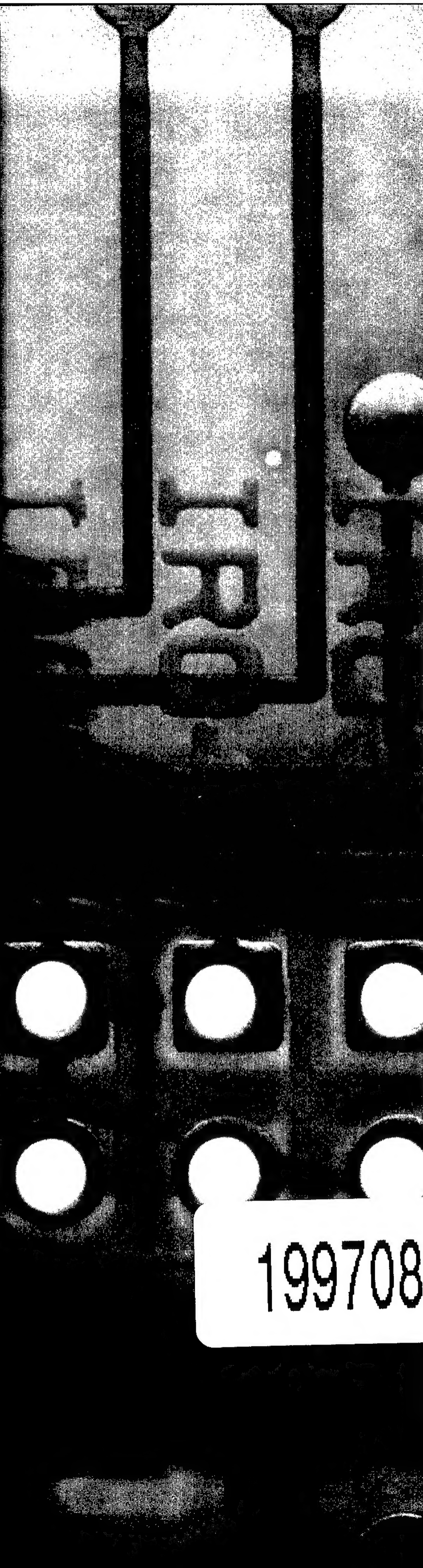

The Journal of

TECHNOLOGY TRANSFER

Volume 22, No. 2
Summer 1997
ISSN 0892-9912

Symposium on Performance Measurement for Public R&D Programs

THE TECHNOLOGY TRANSFER SOCIETY

Published three times a year by the Technology Transfer Society, c/o Bostrum, 435 North Michigan Avenue, Chicago, IL 60611-4067, USA. ISSN 0892-9912. Responsibility for the contents of this *Journal* belongs to the authors, not the Society, its officers, committees or members. For information about the Technology Transfer Society or subscription to the *Journal,* call (312) 644-0828, extension 213, or fax (312) 644-8557.

Technical editing provided by The Institute of Technical and Scientific Communication at James Madison University. For information call (540) 568-8018 or e-mail tsc-program@jmu.edu.

The *Journal of Technology Transfer* is abstracted in the AGRICOLA database, which is accessible through DIALOG, BRS and SilverPlatter CD-ROM. The *Journal* table of contents, Information for Authors and article abstracts are available on Internet through the NASA Mid-Continent Technology Transfer Center's Gopher Server (gopher technology.tamu.edu).

TECHNOLOGY
TRANSFER
SOCIETY

The Journal of
TECHNOLOGY TRANSFER

Vol. 22, No. 2
Summer 1997

SYMPOSIUM
Metrics and Methods for Performance Measurement and Evaluation of Public Research, Technology and Development Programs

Symposium Overview

Gretchen B. Jordan

This collection of papers on metrics and methods for measuring research and technology programs originates from the November 1996 annual conference of the American Evaluation Association (AEA). An international group of people working in the field, anxious to improve the understanding and practice of evaluation and performance measurement of public programs in research, technology and development, have formed a Topical Interest Group (TIG) within the AEA. In addition to encouraging experts from around the world to come together at least once a year to share methods among themselves and with the broader evaluation community, their intention is to establish electronic communication with interested practitioners, share a bibliography that includes unpublished materials, and look for opportunities for communicating best practice in journals, forums, and publications reaching program managers.

This special issue of the *Journal of Technology Transfer* is an opportunity to communicate best practice with an audience that faces great challenges in measurement and evaluation. Science and Technology (S&T) and Research and Development (R&D) programs are required as never before to regularly demonstrate the relevance and value added of their programs. The authors in this publication have practical, hands-on experience with programs all along the research, technology and development continuum, with both planning and implementing measurement and evaluation, and with various metrics and methods.

The papers are rich and contain information and examples previously available only in the gray literature. There is an international perspective, with best practice from the U.S., Canada and the Netherlands, and studies looking at both federal and state government technology programs. Three of the seven papers address the big picture with constructs for measurement and evaluation of research and technology programs. Two discuss evaluations of predictors of program impact, and one explains an innovative case study method. Another analyzes what many state technology programs are doing about performance measurement. With differing levels of detail, all propose metrics for research and technology programs that many reading this issue will find helpful as they struggle with the measurement challenges incumbent with this type of program.

The increasing pressures to demonstrate results and value added are present throughout the world, stemming from general distrust of government and its inability to spend tax dollars wisely in an era when public "wants and needs" appear so much larger than available resources. Whether the distrust is warranted or not, the call for more accountability is loud and clear. The U.S. Government Performance and Results Act of 1993 (GPRA or the Results Act) was modeled after similar requirements in the states and other countries. The Results Act has bipartisan support and Congress is becoming more actively involved as the time of implementation nears. Following a period of pilot projects, the first strategic plans for all but the smallest agencies are due in September 1997. Also due this fall are annual performance plans for the fiscal year 1999 budget. A report on how well the program met the goals in that performance plan and how meeting these has implemented and changed the strategic plan will be due in March 2000.

The challenges of annual performance measurement for research and technology programs are probably familiar to you if you are reading this issue. Research, technology development and the diffusion of that technology can be viewed as a continuum, even though it is well known that there are many feedback loops and very complex relationships among the elements of the continuum. The measurement and evaluation problems of the elements are connected. As each element along the continuum is better able to demonstrate value added, it will strengthen the arguments of the others, particularly if credit is given to the other elements. Research takes credit for research, not for the diffusion of that research, and vice versa.

One major challenge is demonstrating economic and social benefits—the results that legislative bodies are most interested in—such as the number of new jobs created. These benefits only show up when the research or technology has been diffused among some number of the population. The social and economic benefits of research and technology are difficult to quantify because they are often (1) intangible (advance the state of knowledge, changes in behavior), (2) unpredictable (scientific breakthroughs, technology transfer champions), and there is a well-documented (3) complex path and long timelines before outcomes are apparent to the general population.

Value added is multifaceted, however. While many legislators think only in terms of economic and social benefits, there are intermediate outcomes that add value, and a good description of what the program intends to accomplish helps demonstrate the value of a public program. Descriptions of technical programs and what they hope to accomplish, written in a manner that the general non-technical public can understand and relate to, are possible. These descriptions have often been accomplished by example. The Results Act would have them detailed in strategic plans and accompanying performance plans.

The Deputy Director of the U.S. Office of Management and Budget, John Koskinen, has an answer to all who suggest their program is too hard to measure. He suggests they stop doing what they are doing for a few years and see if anyone notices! Rather than do that, the research and

technology communities are investigating how to meet the challenges of measurement. Several U.S. federal inter-agency working groups have been meeting to share methods of measurement. The U.S. Army Research Laboratory (ARL) is a GPRA pilot, showing others an approach to measuring a mission-directed basic research organization, and Ed Brown's paper in this issue describes the ARL evaluation construct. In Canada, there is an integrated approach to performance evaluation that is widely accepted by S&T managers in the government's central agencies. This integrated approach is described in the George Teather and Steve Montague paper. Our paper (Jordan and Mortensen) describes a variation of that approach with examples for both a technology program and a fundamental research program.

These evaluation constructs and the resulting balanced scorecard of performance measures could be very helpful to many readers of this issue. In our experience, many organizations start measuring before they develop a good description of what they are trying to accomplish and how they will achieve their objectives. Without a clear picture of where they are going and what markers to look for along the way, the measures and the units of measurement chosen (metrics) reflect only part of the picture. This is the problem addressed by the balanced scorecard approach to choosing measures across the whole spectrum of performance. Not only do all audiences get the measures they are interested in, such as fiscal soundness and jobs created, but measures reflect intermediate outcomes and customer involvement and satisfaction, so there is an opportunity for continuous improvement.

Another important contribution of this group of papers is the wealth of examples of measures and potential measures of performance for research and technology programs. Rather than reinventing the wheel, organizations can modify what others have done. There is strength in numbers, and the more organizations use similar balanced sets of measures, including intermediate outcomes, the more realistic legislative bodies may be in their expectations for measurement and for program results. The Melkers and Cozzens paper summarizes case studies of what many state technology programs are measuring. The Youtie and Shapira paper shows uses and impacts of a state manufacturing extension program, as the van den Beemt paper shows for a technology grants program in the Netherlands. The Bozeman and Kingsley RVM case study approach traces inputs and intermediate outcomes that can serve as metrics for research and technology programs.

Finally, the underlying theme of this group of papers is that evaluation studies, not just ongoing measurement of specific performance, are necessary if an organization is to have sufficient data to manage its performance. "Metrics" can illuminate situations where performance may not be meeting targets or having the expected results. More in-depth evaluation is needed to find out why and how to improve. More in-depth study is also required to be able to make a cause and effect link between the program's activities and measures of outcomes. The Bozeman and Kingsley paper has a comprehensive table on the applications and strengths of various methods for measuring research and technology programs, as well as the description of their innovative case study approach. There are two good examples of pre and post project evaluations, one involving client surveys, the other expert and staff review. Evaluation studies that document cause and effect and that find predictors of success will strengthen the credibility of using these predictors as performance measures.

Readers will find they have to adapt to the different terminology used by the authors. For example, the Topical Interest Group is named "Research, Technology and Development" Evaluation to allow for the use of "Science and Technology" in some sectors and "Research and Development" in others. The term "Technology Transfer" may not be used explicitly, but transfer of knowledge, technology adoption, and related activities are addressed throughout. Also, because performance measurement is an emerging field, there is no one taxonomy, thus "measure" and "metric" and "indicator" have to be read within the context they are used, and translated to the definitions as the reader uses them.

This is a rapidly evolving area of research evaluation. Hopefully this special issue makes a contribution to the dialogue. Special thanks to John McLaughlin, 1996 Program Chair of the American Evaluation Association Conference for shepherding the offerings of this new Topical Interest Group, and to Robert Hanson of the Social Science and Humanities Research Council of Canada, TIG co-chairman, and to George Teather of the National Research Council of Canada, also a founding father of the interest group. To be part of the topical interest group, or to comment on this issue, contact me at gbjorda@sandia.gov.

Gretchen Jordan
Special Issue Editor
Co-chairman
Research, Technology, and Development Topical Interest Group
American Evaluation Association

Performance Measurement, Management and Reporting for S&T Organizations—An Overview*

George G. Teather

National Research Council of Canada
Ottawa, ON, Canada K1A 0R6

Steve Montague

Performance Management Network
Ottawa, ON, Canada K1P 6A9

Abstract

An integrated approach to performance measurement, management and reporting is presented which builds on the well known logic diagram approach of evaluation theory. The addition of explicit consideration of reach, defined as clients, co-delivery partners and stakeholders, supports a more holistic, balanced approach to the concept of performance, which has found acceptance among S&T performers and central agencies in Canada and the U.S. The description of the "performance framework approach" is supported by rationale for its use at both operational and strategic levels of S&T management. Also included are discussions of recent complementary work and examples of successful use of the approach.

Introduction

There has been a burgeoning interest in the performance of government programs in recent years. This interest comes from several sources, including citizens' concerns about value received for their tax dollars and managers' need to better understand program performance in order to make strategic and operational decisions in an era of declining resources and government expenditure reductions. The Government Performance and Results Act (GPRA) in the U.S. and similar initiatives in other countries reflect this pressure. In Canada, Science and Technology (S&T) has been singled out for improved performance measurement. A major year long review of federal S&T, which began in mid 1994, involved both external consultations with the public, business, universities and other stakeholders as well as an internal review of S&T policies and programs in all science based departments and agencies. The government's response is contained in *Science and Technology for the New Century—A Federal Strategy* (Canada 1996), which includes a commitment to the assessment of federal S&T performance on a regular basis, exemplified in the following quotation: "Each department and agency will set S&T targets and objectives, establish . . . performance indicators . . .". In order to respond to these challenges, the S&T community in these and other countries is under considerable pressure to develop mechanisms to determine and measure performance in a credible, logical manner which will be understood by the government and other key stakeholders.

In recent years there have been numerous efforts to measure S&T performance with many examples of good studies; unfortunately, they are interspersed with poor ones. Using an analogy borrowed from the technology sphere, S&T performance measurement is still an emerging capability, on the initial slope of the "S" curve, characterized by many competing initiatives with varying degrees of quality and capability, each striving for acceptance and survival.

Drawing on over ten years of experience in evaluation of government S&T organizations and programs, the authors have developed an integrated approach to the consideration of S&T performance which has found acceptance by S&T managers and government central agencies in Canada and the U.S. In this paper, we will link this approach with a number of other recent initiatives and complementary advances, some of which reside in the gray literature of government reports. The intention is to present a comprehensive, coherent framework for understanding and describing the role of S&T in the modern economy, a necessary precursor to measuring, managing and reporting on the performance of individual organizations or programs.

*The authors acknowledge the contributions made by many colleagues to the development and refinement of the performance framework approach for use in S&T. In particular we wish to thank Robert McDonald of Industry Canada, Marielle Piché of the National Research Council of Canada, Aileen Shaw of the Canadian Space Agency, and Gretchen Jordan of Sandia Laboratories, United States Department of Energy for their contributions. Their support for the performance framework concept has provided valuable experience in the use and benefits of the approach.

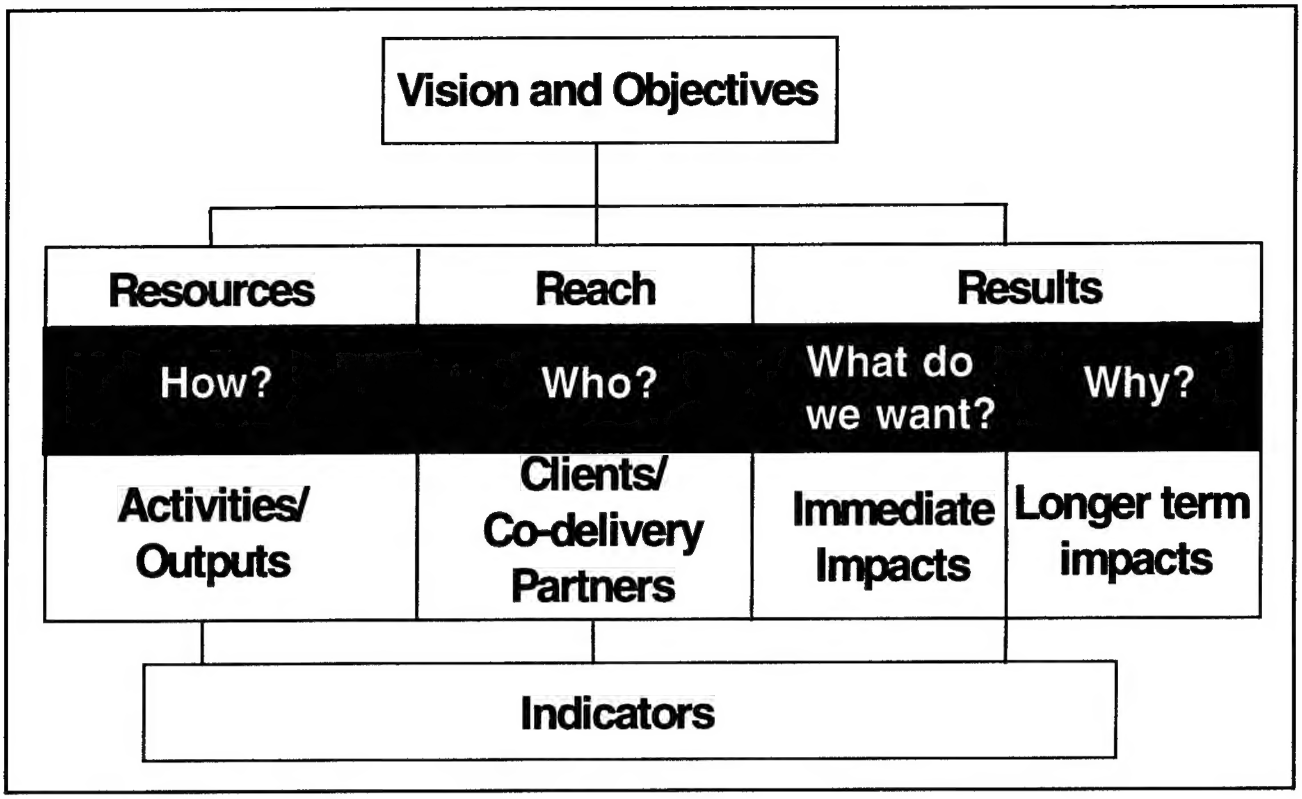

Figure 1. Performance framework

The Performance Framework Approach

In the late 1970s, the Canadian federal government institutionalized the use of the logic model originally introduced by Joe Wholey and others (Wholey 1980) as a basic tool for evaluation of federal programs. Consequently, there has been extensive experience in the use of the logic model since that time. The performance framework (Montague 1993; Montague 1994) (Figure 1) was developed from the logic model which was modified to include explicitly consideration of the "reach" of the program or organization under review. Reach defines the target clients, key co-delivery partners and stakeholders which are the mechanism through which activities and outputs are transformed into results. Rather than focus on impact, this approach considers performance in terms of the entire program in a holistic manner, linking resources to reach and results. This performance framework is congruent with Kaplan and Norton's "balanced scorecard" approach (Kaplan et al. 1996) of business management theory as they each describe successful performance in terms of a spectrum of factors internal and external to the organization which relate to both capability and results.

Explicit consideration of reach is perhaps the most novel element in this approach. The pathway between activities, outputs and results always includes transfering knowledge or technology to another person or organization, and therefore involves "reaching" someone else outside the organization creating the S&T output. Inclusion of reach considerations in policy and program design, planning and performance analysis forces consideration of the receptor population and whether there is receptor capacity for a given S&T initiative.

In addition to providing a rational approach to understanding the linkages between resource utilization, resulting capability and consequential results, the performance framework focuses directly on management needs by responding to stakeholders' key questions in a straightforward manner. The questions How? Who? What do we want? and Why? can be answered directly using this approach. Some may be concerned that the causal linkage of activities and outputs to objectives is intended to defend the status quo; in fact, the opposite is true. The performance framework challenges both existing strategies and their operational implementation to demonstrate performance against objectives or alternately provides a mechanism to consider alternate service delivery approaches in terms of better defined performance objectives..

Figure 1 captures the key attributes of the performance framework approach. Conceptually, resources (staff and operating funds) are used to perform activities and create outputs. This is **HOW** one goes about achieving objectives. These activities and outputs reach a target client group either directly or with the aid of co-delivery partners and stakeholders. This is **WHO** is affected by the activities and outputs. As a result of the activities and outputs, the target client group behaves differently, and immediate impacts occur. This is **WHAT** happens. Over

the longer term, the changed behavior leads to more extensive and consequential impacts. If the program is performing well, these changes can be causally linked to intended long-term program objectives. This responds to **WHY**. Sources of information to measure program performance can then be identified, and performance indicators can be developed in terms of these themes for any given program or organization.

This performance framework approach has been used successfully to describe the performance of many programs in a wide variety of disciplines. It has been adapted for use in the S&T domain by the authors and colleagues and linked to other initiatives focusing on S&T policy and impact measurement methodologies. This integrated systematic approach has received broad acceptance from a number of S&T managers and stakeholders at the program and organizational level in both Canada and the United States.

Rationale for Government S&T

The performance framework approach does not determine program objectives, but rather adopts those objectives which have been developed through policy or program decisions by government or senior management to describe performance. Consequently, the rationale and policies defining the role and purpose of government S&T need to be articulated within the performance framework context to define intended long term impacts. Several recent initiatives have helped to better understand and describe the role of government S&T in modern society.

Economists traditionally have used the non-appropriable externalities or *"public good"* nature of much S&T activity to explain under investment by the private sector and the need for significant government investment in S&T, basic R&D, standards, and related work (Arrow 1962). However, as well as investing in S&T on behalf of the private sector, the government is also a major user of S&T to make and implement policy and regulatory decisions to define and manage the society in which we live. Because of the generic nature of much S&T knowledge, it provides a foundation for use by both the public and private sectors and supports national competitiveness as defined broadly to include a well educated, healthy population, and effectively operating society underpinning the efficient production of goods and services which are competitive in price and quality.

This concept was carried further by Greg Tassey, an economist with the National Institute of Standards and Technology (NIST). Tassey notes the existence of technology infrastructure as a key to economic development in *Technology Infrastructure and Competitive Position* (Tassey 1992). Technology infrastructure comprises an economy's set of institutions and facilities relating to its science base, generic technologies, applied technologies, and *"infratechnologies,"* that is, technical *"tools"* such as test methods and measurement techniques or protocols that affect the productivity of research and the diffusion of innovation.

As mentioned previously, there has been extensive examination of the role of the government as an investor in S&T to benefit the private sector, but rarely as a consumer using S&T to meet its internal needs. As well as the obvious role of S&T in defense and public health, government S&T organizations and resources have contributed to the achievement of government objectives in the areas of agriculture, the environment and construction, to name just a few. A Canadian study (Canada 1993) on socioeconomic impacts of government S&T found that S&T was performed broadly speaking for four purposes: building of S&T competence, policy development, policy implementation and industrial development. Much recent emphasis has been on this last category and the impact of government S&T on direct wealth creation. This focus is also discussed in the recent examination of the role of government laboratories in the U.S. by Papadakis (1995).

In fact, evaluators and analysts of S&T programs have come to recognize that *"innovation"*—the essential core product of S&T—affects behaviors across a wide range of institutional actors in both public and private sectors. The influence cannot and should not be constrained by simply analyzing private or even narrowly defined social returns on investment (Lipsey et al. 1996).

Application of Performance Framework

The use of a performance framework model to respond to How? Who? What do we want? and Why? facilitates an analysis of the behavior changes and benefits that occur within major institutional actors as a result of S&T and related activities.

As an example, imagine the development of new software which results in greatly improved images from remote sensing satellites. The private benefits stream of this innovation may be minimal, as very few direct jobs or sales are created in the software firm developing the product. After all, new software requires none of the production *"gear-up"* that would accompany a machinery innovation. With competition in this field and difficulties in intellectual property protection, imitators may soon in fact erode any private competitive advantage for the developing firm.

But consider the broader behavioral effects on users of data from satellites resulting from this innovation. With more precise and reliable information available, the ability to make natural resource allocation decisions relating to agriculture, forestry and environmental protection is improved. Emergency response to natural phenomena such as landslides, storms, forest fires, and oil spills can be better managed. In the longer term, more exacting mapping standards may emerge leading to the development of world-class expertise in the field which in turn generates

HOW?	WHO? WHERE?	WHAT do we want?	WHY?
Fundamental research	Science community	Advance knowledge	Wealth creation, public health, security, and environmental protection
Applied research, development, and technology transfer support	Specific public and private users	Technology adaptation, adoption, development, and exploitation (in support of public missions as well as private benefits)	
Innovation system support	Industry groups/ sectors and consumers	Improved innovation speed and efficiency and reduced market transaction costs	

Figure 2. S&T performance framework

spinoffs in scientific equipment, consulting, and various natural resource management services. All of these benefits may accrue from as little as one software innovation in the right place at the right time.

The performance framework leads the analysis beyond the natural tendency to focus on immediate direct impacts of each innovation (e.g., product sales) to an examination of a broad range of benefit streams. These include behavioral changes beyond the advancement of knowledge and the adoption of technology by specific users to innovation *"system"* effects relating to large institutions, standards, and related sectors of the economy. Figure 2 shows a general application of the framework to the S&T domain.

Performance Measurement—Practice

As well as defining objectives, use of the performance framework approach requires the collection and analysis of performance based information in terms of the categories of resources, reach, and results.

In most cases, information on resources is relatively easy to obtain, since program management and information systems have traditionally focused on resource utilization. Budget allocation, categories of staff and outputs such as papers and reports published, seminars held etc. have been readily available and extensively used as a proxy for impact and overall performance in the past. However, a refereed publication, although a legitimate indicator of productivity and quality, has no impact outside the laboratory which produced it until and unless someone else does something different than they would have without having read the article or heard about it at a conference or seminar (citation is a legitimate indicator of impact).

Reach needs to be understood conceptually as a precursor to data collection and analysis. Reach can include many groups, the first being target and actual clients or recipients of the outputs. Another could be those with complementary skills which, if induced to participate, can increase the likelihood of achieving positive results dramatically. An example from recent experience is the increased linkages between researchers and technology transfer specialists in universities or government laboratories, which have been found to increase the successful transfer and utilization of S&T outputs significantly. A third group is key stakeholders, who can provide credibility and support. An example would be an industry association representing the target client group whose support might induce members of the target client group to become clients. The last major category of reach to be considered is the beneficiaries of the S&T activities beyond the direct clients. For demonstration projects with one firm, this could be the larger industrial sector targeted as potentially utilizing the technology.

Information on various aspects of reach has often not been previously considered as necessary and is therefore not typically available. For example, performance-related analysis such as penetration of intended target client groups can be problematic. For some S&T programs, target client groups or recipients of outputs have not been fully identified. Targets can be as broad as the international R&D community or as narrow as a single private firm within

Table 1. Methods useful for assessment of past R&D

	R&D Purpose			
R&D Type	Category 1 R&D Infrastructure	Category 2 Policy Development	Category 3 Policy Attainment	Category 4 Industrial Development
Basic/Strategic	(Modified Peer) (Partial Indicators)	Modified Peer (Partial Indicators)	Modified Peer (Partial Indicators)	Modified Peer (Partial Indicators)
Applied	(Modified Peer) (Case Studies) (Partial Indicators)	Modified Peer User Surveys Case Studies (Benefit-Cost) (Partial Indicators)	Modified Peer User Surveys Case Studies (Benefit-Cost) (Partial Indicators)	Modified Peer User Surveys Benefit-Cost Case Studies (Partial Indicators)
Development	(Modified Peer) (Case Studies) (Partial Indicators)	Modified Peer User Surveys Case Studies (Benefit-Cost) (Partial Indicators)	Modified Peer User Surveys Case Studies (Benefit-Cost) (Partial Indicators)	Modified Peer User Surveys Benefit-Cost Case Studies (Partial Indicators)

*Use of brackets signifies potential for use in particular circumstances.

industrial sectors (i.e., pharmaceuticals) or government policy groups responsible for regulation as examples of intermediate level targets. Often information systems do not capture client information, and performance analysis in terms of reach (penetration of target client group) is difficult to perform. Reach is defined to include co-delivery partners. For many government S&T programs, effective linkages with private sector partners or industry associations can have a major influence on achievement of results. In the case of NIST and Canada's NRC, private sector calibration laboratories are an important means to reach the intended audience of producers and users of measurement equipment.

Results, defined as "What do we want?" and "Why?", are particularly difficult to measure for many S&T activities. The long pathway between S&T and ultimate impact, with the many intervening factors which come in to play, including business cycles, interest rates, politics, etc., can make attribution and causality difficult to determine. Immediate impacts are usually more directly attributable to the S&T whereas, except in certain cases with few intervening factors, longer term impacts become more difficult to claim. In practice, even immediate impacts are often difficult to determine, especially if client information is not kept, since it is usually necessary to have some indication of the change in client behavior to assign impacts. Often a combination of quantitative and qualitative information on service standards, client awareness and use of S&T can be collected by using client surveys, end-of-project feedback forms, and file analysis. In the authors' experience, many S&T programs, while ignoring immediate impacts, attempt to determine results in terms of longer term impacts in spite of the difficulties, as a response to the need for accountability and continued funding. A more balanced approach to measurement, capturing indicators of both immediate and longer term results is usually more useful to program management and credible to stakeholders. While care needs to be taken to keep performance measurement efficient, expanding the utilization and resulting benefits can compensate to some extent.

There are a number of reports and books which identify and describe methods for determining the immediate and longer term results or impacts of S&T. Some are quite technical, as they are written for an expert audience. One review intended for non-specialists, mentioned previously, is a study entitled *Methods for Assessing the Socioeconomic Impacts of Government S&T* (Canada 1993), which describes and analyses the major methods available and their applicability and provides an extensive bibliography of published and gray literature from various countries. Table 1, from that report, presents a summary of the applicability of various methodologies for R&D performed for various purposes. In this table, traditional peer review has been modified to include greater input on the potential of downstream utilization of research to complement the focus on research quality. The partial indicators identified in this table are closely related to the general performance framework approach being discussed, requiring the identification of a number of types of information, each of which provides a partial indicator of impact. Following the approach identified in this paper and making use of the appropriate methodologies in Table 1, it should be possible to use several complementary methods to perform a credible assessment of the performance of virtually any S&T program.

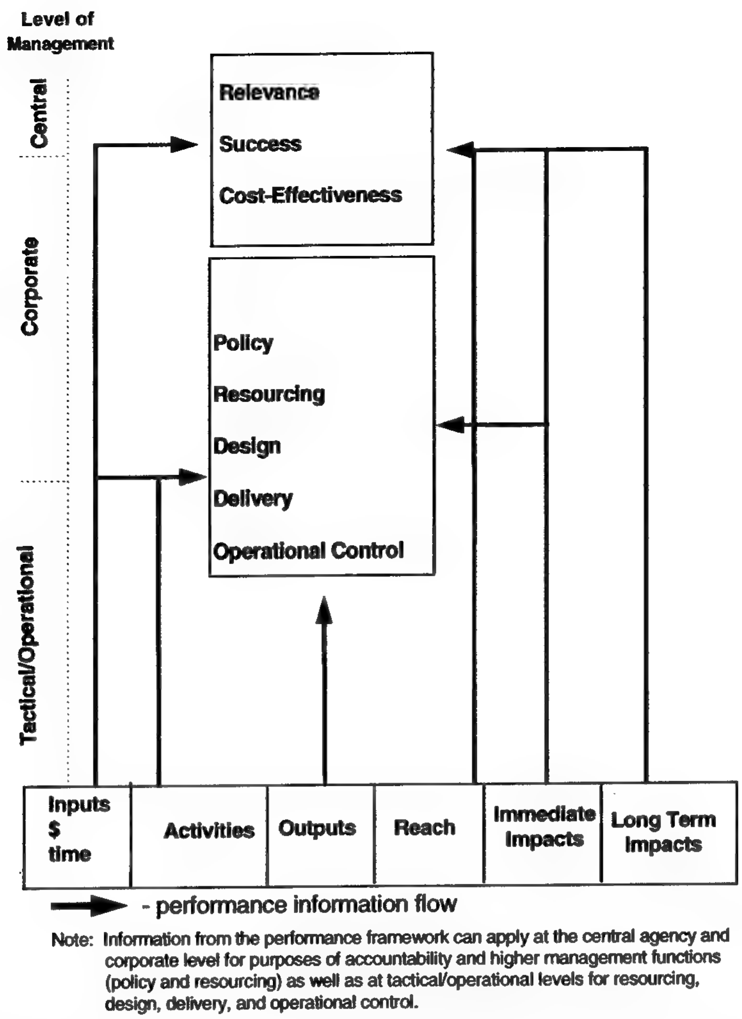

Figure 3. The relationship of key performance information to different management levels

Examples of the Successful Use of the Performance Framework

The basic elements of the performance framework approach have been used by the authors since the late 1980s in assessment work at the National Research Council of Canada and other S&T organizations in Canada. The 1990 Industrial Research Assistance Program (IRAP) Evaluation Study (Canada 1990) used the basic performance framework approach, examining resources, reach and results. The study included extensive analysis of the penetration of IRAP into the Canadian Manufacturing Sector as well as the immediate and longer term impacts of IRAP assistance as reported by assisted firms. IRAP, a technology extension program, was found to be highly incremental, and clients attributed a considerable share of their success to IRAP assistance. This assessment was used as a reference document by a 1991 Parliamentary Inquiry into the program as an important input to decision making and was quoted extensively as the basis for conclusions and recommendations.

Another extensive assessment of the same program, documented in *Assessment of Industrial Research Assistance Program—Review Committee Report* (Canada 1996), has just been completed using a similar performance framework approach which provided a comparative analysis of intended and unintended changes to the program five years later. The Review Committee responsible for the assessment, made up of program stakeholders external to NRC, reported that the performance framework approach was an effective method to collect credible evidence on the overall performance of IRAP and to develop recommendations on key aspects of IRAP as input to a new Strategic Plan for the next five years.

There are many other examples of successful use. The Canadian federal industry department, Industry Canada, has developed a guide to assist managers in understanding and measuring performance (Canada 1995), and the Canadian Technology Network (CTN), a recent initiative of the federal government, adopted the framework approach to assist with monitoring and managing both implementation and ongoing network performance. The document *An Evaluation/Performance Framework for the Canadian*

Technology Network (Canada 1995) contains an extensive description of the principles of the performance framework as well as a practical example of the use of those principles to determine key performance characteristics for CTN.

The performance framework approach is relevant to many levels of management and S&T decision making. Figure 3, reproduced from the CTN study (Canada 1995), demonstrates the relevance of performance information to various levels of management. Operationally, attention is primarily focused on resource management and delivery—with some reference to reach and immediate impacts. As the focus changes from program delivery to strategic and corporate to government level considerations, there is progressively more attention paid to longer term impacts. For major S&T organizations and at the national level, there is a clear requirement to link program impacts to government S&T policy objectives.

As a result of the recent federal government S&T review in *Science and Technology for the New Century—A Federal Strategy* (Canada 1996), Industry Canada (similar in many aspects to the U.S. Department of Commerce) has embarked on a new approach to corporate governance and policy analysis for S&T. The goal is to determine the effectiveness of policy initiatives in terms of performance, according to the *Science and Technology for the New Century—Industry Canada's Action Plan* (Canada 1996). While in theory this was always the objective, increased attention to the collection and utilization of credible information linked to the performance and effectiveness of specific policy initiatives will support improved implementation of policy decisions as well as promote more informed choices among policy alternatives.

Conclusions

Initial experience in the application of a performance framework for the analysis of S&T performance has been promising. Frameworks developed for specific programs and organizations have been shown to assist S&T performance planning, measurement, and reporting. The approach helps resolve traditional conceptual difficulties such as inappropriate narrow considerations of benefits and impacts, and provides a practical, consistent template for information collection, analysis, and reporting on performance.

References

Arrow, Kenneth. "Economic Welfare and the Allocation of Resources for Invention." In *The Rate and Direction of Economic Activity: Economic and Social Factors*. Princeton: NEBR, 1962.

Canada. Federal Government. *Methods for Assessing the Socioeconomic Impacts of Government S&T.* May 1993.

Canada. Industry Canada. *Focusing on Results: A Guide to Performance Measurement.* March 1995.

Canada. Minister of Supply and Services. *Science and Technology for the New Century —A Federal Strategy.* 1996.

Canada. Minister of Supply and Services. *Science and Technology for the New Century—Industry Portfolio's Action Plan.* 1996.

Canada. National Research Council. *An Evaluation/Performance Framework for the Canadian Technology Network.* August 1995.

Canada. National Research Council. *Assessment of Industrial Research Assistance Program—Review Committee Report.* November 1996.

Canada. National Research Council. *Industrial Research Assistance Program—Evaluation Study Final Report.* October 1990.

Kaplan, Robert, and David Norton. "Using the Balanced Scorecard as a Strategic Management System." *Harvard Business Review,* January-February 1996, pp. 75-85.

Lipsey, Richard G., and Ken Carlaw. "A Structuralist View of Innovation Policy." In *The Implications of Knowledge-Based Growth for Micro-Economic Policies*, ed. Peter Howitt, Ministry of Supply and Services. Canada: University of Calgary Press, 1996.

Montague, Steve. "The Three Rs of Performance-Based Management." *Focus,* December/January 1994, pp. 26-28.

Montague, Steve. *Performance Framework Approach to Public Management.* January 1993.

Papadakis, Maria. "Federal Laboratory Missions, Products and Competitiveness." *Journal of Technology Transfer* 20, April 1995, pp. 54-67.

Tassey, Greg. *Technology Infrastructure and Competitive Position.* Norwell, MA: Kluwer, 1992.

Wholey, Joseph. *Planning Useful Evaluation: Evaluability Assessment.* Beverly Hills, CA: Sage, 1980.

Author Biographies

George Teather has been conducting evaluations for the National Research Council of Canada for 12 years, examining a wide range of S&T activities, including research and development, library services, information dissemination and technology assistance to firms. Before

becoming an evaluator, Mr. Teather co-authored a number of scientific articles as a researcher in NRC's laboratories. He is a professional engineer with a Masters degree in Electrical Engineering and is at present the Chair of the professional development committee of the Canadian Evaluation Society.

Steve Montague is a Director of Performance Management Network, a management consulting firm focusing on performance related issues. Mr. Montague has been involved in supporting management decision making in public and private sector organizations in S&T and other sectors for over 15 years. Over the past several years he has delivered many workshops and presentations on performance based management, including recently a number for senior Canadian federal government managers. Mr. Montague has a Masters degree in Public Administration.

Measuring the Performance of Research and Technology Programs: A Balanced Scorecard Approach*

Gretchen B. Jordan
Sandia National Laboratories
Arlington, VA 22209

John C. Mortensen
Sandia National Laboratories
Arlington, VA 22209

Abstract

Research programs, like other government programs, are now being requested to demonstrate relevance and value added to national social and economic needs. Complexity, unpredictability and other factors make it difficult to define specific performance measures for R&D programs. This paper describes the performance measurement efforts of one technology development program within the U.S. Department of Energy and proposes a strategy for applying this balanced scorecard approach to a fundamental research organization. Simple logical models of the inputs, activities, outcomes and long term results of R&D programs are proposed. A critical few measures of performance that answer questions from multiple audiences are then chosen across this performance spectrum.

The Challenge

There has been a growing interest around the world in measuring performance and "managing for results" in the public and private sectors, largely driven by increasing competition for limited resources and increasing calls for accountability. Performance management, the systematic development and application of performance information to plan and continuously improve programs, is mandated for the U.S. federal government through a series of laws and executive orders. These include the Government Performance and Results Act of 1993, Government Management Act of 1994, Chief Financial Officers Act of 1990, Federal Acquisition Streamlining Act of 1994, Information Technology Management Reform Act of 1996, Executive Order 12862 Setting Customer Service Standards of 1993, and the National Performance Review (U.S. DOE 1996).

*This paper describes work performed by Sandia National Laboratories Energy Policy and Planning Department, Albuquerque, New Mexico 87185, with the support of the U.S. Department of Energy under contract DE-AC04-94AL85000. The continued support and enthusiasm of Darrell Beschen in the DOE Office of Energy Efficiency and Renewable Energy and Iran Thomas in DOE Office of Energy Research, Basic Energy Sciences, is appreciated. The authors also acknowledge the contributions of John Reed and the training in the logic chart and performance spectrum provided by Ron Corbeil and Steve Montague. The opinions expressed are those of the authors and do not represent the opinions of the U.S. Department of Energy.

Research programs, like other government programs, are now being requested to demonstrate relevance and value added to national social and economic needs such as education, economic competitiveness, the environment, and national security. There are several challenges in defining specific results-oriented performance measures for R&D programs, particularly the annual quantitative measures requested by legislative bodies. Technical programs have difficulty communicating relevance and value added with non-technical audiences because they have not typically had to do so. The customers for research products are usually other technical organizations, with taxpayers the ultimate, but indirect, customer. Also, the process from research and technology development activities to noticeable societal impacts is often complex and occurs over a long period of time. This complexity makes demonstrating cause and effect very difficult. Moreover, the results of R&D are unpredictable. Breakthroughs are sometimes serendipitous, and it is not unusual for a scientific breakthrough to languish on the shelf until a change in circumstance reveals a useful application in a technology market. Even after an application is determined, the diffusion of a technology into the marketplace still depends on many variables outside the control of the program.

This paper describes the performance management approach developed by Sandia National Laboratories in collaboration with a technology development program within the U.S. Department of Energy (DOE), and proposes a strategy for applying that approach to a fundamental research organization. It builds on a body of literature that is growing as the research community addresses these new requirements. The U.S. National Science and Technology Council provides examples of current approaches in

its report *Assessing Fundamental Science* (U.S. Science and Technology Council 1996). Examples include the U.S. Department of Energy performance based contract assessment methodology and an evaluation model developed by the federal Interagency Research Roundtable. In the area of applied technology programs, comprehensive discussions of measurement can be found in a collaborative effort by the Institute for Industrial Research (Tipping et al. 1995) and in an annotated bibliography produced by the International Center for Research on the Management of Technology (Hauser 1996).

The Balanced Scorecard Approach

In order for a program to reach its strategic goals, manage its performance, and be accountable, it must be able to collaboratively identify, define, and perform a spectrum of essential functions. Basic elements of a performance spectrum are inputs, activities, outputs, outcomes, and long-term results. At every point within this spectrum, performers of activities must interact with customers who are or will be users of the products and services, partners who help to produce or deliver the products and services, and stakeholders who impact or are impacted by the organization.

Different audiences have questions about different elements of the performance spectrum. Those interested in long term outcomes ask, "What is the relevance of the program, that is, who has a need that is filled by it?" Other audiences ask different questions. Some are more interested in short term results and ask, "Is the program effective, having results, successful?" Others want to know about resource management practices and ask, "Is the program efficient and well managed?" Others want to know about the people and organizations involved. Finally, just as the private sector asks about return on investment, public programs are often asked to provide a measure of cost effectiveness, that is, a ratio of the resources expended to the results achieved.

Choosing a "balanced scorecard" of measures means choosing measures that reflect each element of the performance spectrum as well as combinations of elements. The term has been popularized by articles in the Harvard Business Review describing an approach that also develops a balanced set of performance measures (Kaplan and Norton 1996). The approach provides a balanced picture of the health of the organization that will satisfy all the key audiences. For example, a measure of an input might be total dollars available to a program. An output measure might be units of product delivered. These two combined form an efficiency measure—the cost per unit of production. The balanced scorecard approach also helps ensure that the measures chosen will drive performance toward stated program goals. It recognizes that a tradeoff exists among the areas of resources management, reach to targeted populations, and results. By measuring in all three areas, the possibility that measurement concentrates on one area to the exclusion of another and thus has perverse effects is minimized. The research community, for example, is concerned about annual performance measurement causing undue emphasis on short term gains at the expense of projects that have less obvious short term benefits but potentially large rewards in the long term. And the technology transfer community may be able to stimulate more energy savings by focusing on large companies, but may also have a mandate to work with small businesses.

There are other benefits to the balanced scorecard approach. In addition to helping an organization describe a shared vision of the full spectrum of its performance, the approach enables the organization to measure and evaluate as it progresses for continuous improvement. Of course, a balanced performance story has to be developed through collaboration, and this collaboration in itself has benefits such as improved communication among organizational levels. However, with the benefits of the balanced scorecard approach come certain prerequisites. The approach does not succeed without an accompanying performance management infrastructure, including a corporate framework for using and improving performance data, performance incentives aligned with the measurement system, and the knowledge, skills and tools to implement the system.

Performance Planning Tools for Developing the Scorecard

Our balanced scorecard approach is called COREporate™ Performance Planning because the process helps describe performance for a "corporate" level of organization as well as for individual programs. The two tools used for developing a picture of the organization's performance are the performance spectrum described earlier and the logic chart. Both have been used extensively by Canadian evaluators (Corbeil 1992; Montague 1993; Montague 1996), and the logic chart evolved from Wholey's evaluability assessment work (Wholey 1980). The logic chart captures the logical flow and linkages that exist in any performance spectrum and is used to organize and simplify the performance spectrum. In addition to viewing programs within organizations, the tools view annual progress goals as these relate to multi-year goals and the budget.

In a collaborative, iterative way, the picture of the program or organization's desired results, and the resources and path to achieving those results, is developed. At the end of the process all stakeholders have a shared view of performance expectations for the program or organization. Both the performance spectrum (in table form) and the logic chart diagram are used in performance planning sessions. Although the logic chart is primarily used to describe programmatic activities, we have extended the diagram to include, in list form, the management of resources supporting the program and the customer groups

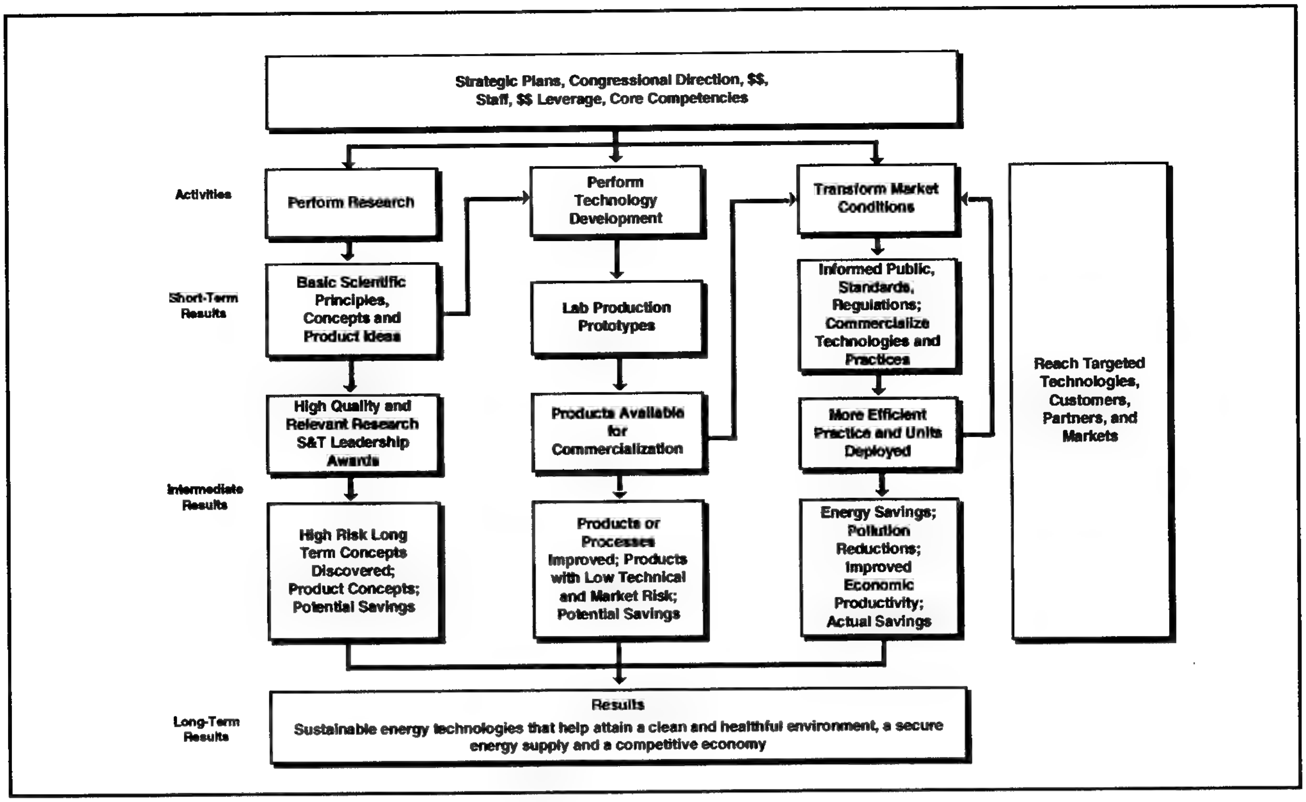

Figure 1. An EE logic chart for R&D activities

reached. Figure 1 is a logic chart for a technology development and deployment program. Figure 4 is a logic chart for fundamental research.

An Example: The Technology Development and Deployment Performance Story

For the past four years, Sandia National Laboratories has assisted the Office of Planning and Assessment (OPA), which reports directly to the DOE Assistant Secretary for Energy Efficiency and Renewable Energy (EE), with performance measurement and evaluation activities. Using the logic chart and the performance spectrum, we have helped EE programs complete performance plans and choose balanced sets of performance measures that can cascade throughout an organization. Corporate performance planning has been done with the DOE Climate Change Action Plan programs (Jordan and Beschen 1995), the Federal Energy Management Program, and for two annual calls for EE-wide performance data. In all cases, the purpose of the performance information was to report progress and to gather data to improve programs. Collaboration among program managers, attention to measures expected by the various audiences, and use of these performance planning tools has led EE to a performance story and a balanced set of measures that may be useful to other technology development and deployment programs and perhaps are generic enough to be useful to other types of organizations.

The U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy collaborates with scientists, consumers, suppliers, financiers, regulators, and other government organizations. It performs research, develops new and improved products and processes, and provides policy, standards, technical tools, and information that will accelerate and expand the adoption of energy efficient and renewable energy technologies. The adoption of these technologies will result in energy and energy cost savings and increased use of alternative energy sources such as wind and solar thereby displacing fossil fuels. In turn, this will result in a cleaner, healthier environment, less dependence on imported oil, and opportunities to invest private and public cost savings to meet other objectives. The office is structured around the end-use sectors for which its technologies are being developed: buildings, industry, transportation, utilities, and the federal government. EE received $845 million in fiscal year 1996.

The EE Performance Spectrum in a Logic Chart

Just as the description above captures the essence of what EE does and the expected long term results, the logic chart in Figure 1 graphically displays the same information. Reading from left to right just below the inputs in the chart, EE activities (in a simplification of a non-linear process) are to perform applied research, develop technologies, and deploy technologies and transform markets. In the rows below these activities, the chart shows the result-

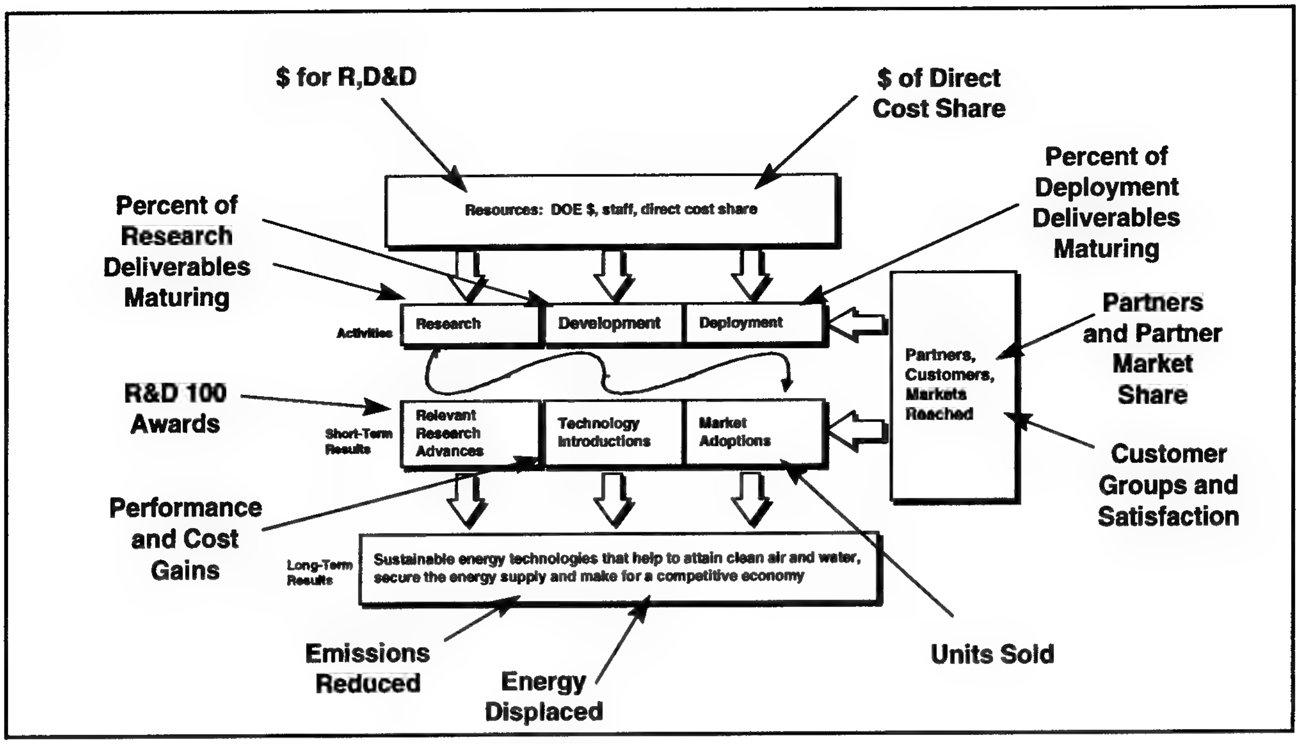

Figure 2. EE's measures are balanced across the performance spectrum

ing outputs and progress or outcomes from each of EE's primary activities. As shown later, the sequence of events under EE's "perform research" track is similar to the more detailed logical performance path proposed for fundamental research. An example of a path from activities to results can be seen for transforming markets. In Figure 1, notice that transforming markets through policy and procurement actions, regulations and standards, as well as providing technical tools, advice and information dissemination, leads to a public that is informed about the technologies and motivated or required to acquire them. The adoption and use of the technologies leads to actual energy savings, pollution reduction, and accompanying benefits.

That EE is reaching the right people is indicated in the box on the righthand side. EE can describe the major partner and customer groups that it works with, as was done in the introduction to the EE program above. Individual programs can be more specific, and the use of shared definitions of categories and characteristics of partners and customers allows for summary analysis and comparison across programs of data such as customer satisfaction.

The resources that EE manages for its programmatic activities are alluded to in the box at the top: EE staff, core competencies and funds, along with leveraged funds, implement strategic plans and Congressional direction. As mentioned earlier, it is necessary to measure resource management even as one moves to more "results" oriented measures, both as a stand alone measure of fiscal responsibility and as part of a cost effectiveness ratio.

Balanced Measures for a Technology Development and Deployment Program

After an organization's performance spectrum is defined, choices of essential and balanced measures become clear. These are then improved through collaboration and experience. The set of corporate measures that has evolved over time is shown in Figure 2 overlaid on a simplified EE logic chart. Corporate measures for resource management include: DOE funding by activity type, cost share dollars, and critical deliverables maturing (also outcomes measures); for people reached: partners and the market they represent, and customer groups reached and satisfied; and for short and longer term results: R&D awards, technology cost and performance gains, units deployed, energy displaced (saved or displacing fossil fuels, in trillion Btus) and greenhouse gas emissions reduced (in million metric tons of carbon equivalent).

The EE critical deliverables maturing measure was added in 1996 as a means of aggregating the short and intermediate term performance of all EE programs and linking measurement to the budget. The metric is composed of measurable targets for each deliverable, grouped by activity or strategy, and accompanied by the cumulative DOE dollars required to achieve the deliverable. An example is shown in Figure 3. The DOE dollars within and across the programs that are spent on all these deliverables by year are aggregated across all the activities or strategies. The metric quantifies how much R&D and Deployment is being completed on time and within budget on an annual basis. The intention is to be able to report, for example, that in the past year $400

Example: Strategy to Integrate Vehicle Systems			
Year	Target	Deliverable	Cumulative Cost of Deliverables (DOE $)
[Example] 1996	[Example] 3 subcontracts	[Example] Contract initiated with Chrysler	[Example] $0.5 Million
1997	50 mpg five passenger vehicle	Complete fabrication of 50 mpg vehicle	$17.0 Million
1997	50 mpg five passenger vehicle	Test vehicle to finalize operations strategy and package design	$16.0 Million
1998	Complete development of hybrid propulsion systems	Complete Ford and GM hybrid propulsion system and integrate into current year production vehicle	$20.0 Million

Figure 3. Supporting information for "Critical Deliverables Maturing" metric

million in research deliverables were delivered, which represents 80% of the $500 million in research deliverables that had been projected for the year and 45% of EE's budget (the data used is illustrative). Corporate tracking would inform the Assistant Secretary which R&D and Deployment areas are delayed or over budget so that these problems might be addressed. Dollar values serve as a proxy for the significance of different deliverables, and, of course, link performance with the budget. This is far from a perfect weighting scheme but recognizes that not all deliverables should be treated equally.

Applying the Performance Spectrum to Fundamental Research

The lessons learned in describing and choosing measures for a technology development and deployment organization can be applied to fundamental research programs. The performance story of fundamental research can be told in words and measures that those outside the research community can relate to and understand. Like the technology development and deployment story, the fundamental research story will describe how resources are directed toward research in specific areas, based upon strategic plans developed knowing the program's mission, core competencies, and the historical context of problems and significance of and potential for solutions. It will describe how resources are managed and activities are monitored to ensure quality. It will describe the progression from activities to outcomes, such as new knowledge, some of which may have known potential impacts. Further, that progression continues when the new knowledge finds application in more applied research or technology development and deployment. The story of fundamental research will also describe the players involved. Who is doing this and related work? Are the right partners involved, are useful products provided to sponsors and other researchers, and eventually to technology developers and the taxpayers?

Building the Performance Story

The first step in building a performance story that addresses the full spectrum of program performance is to answer the basic questions regarding results, management of resources, and people reached. The second is to organize our answers in a logical fashion across the performance spectrum, perhaps displaying this in a logic chart. The third step is to choose a balanced set of measures. These steps are completed below for a fundamental research program. Parts of the story were revealed in a performance planning session held at the federal Interagency Research Roundtable. Parts are based on experience working with the DOE Office of Energy Research, in particular with the Office of Basic Energy Sciences.

Step 1. Describing the performance spectrum.

What will be our long-term results? State a clear objective for the program. For example, a DOE fundamental research program has the strategic goal of "better understanding through new insights and knowledge into the nature of energy and matter." Since this is mission-directed research it is also supposed that the knowledge gained will eventually be useful in more applied research and lead to development of new and improved sources of energy that have positive economic impact. While applications of research are not under the control of the program, it is important for the research program to state the anticipated link and perhaps track applications that have occurred in order to demonstrate the value of past research.

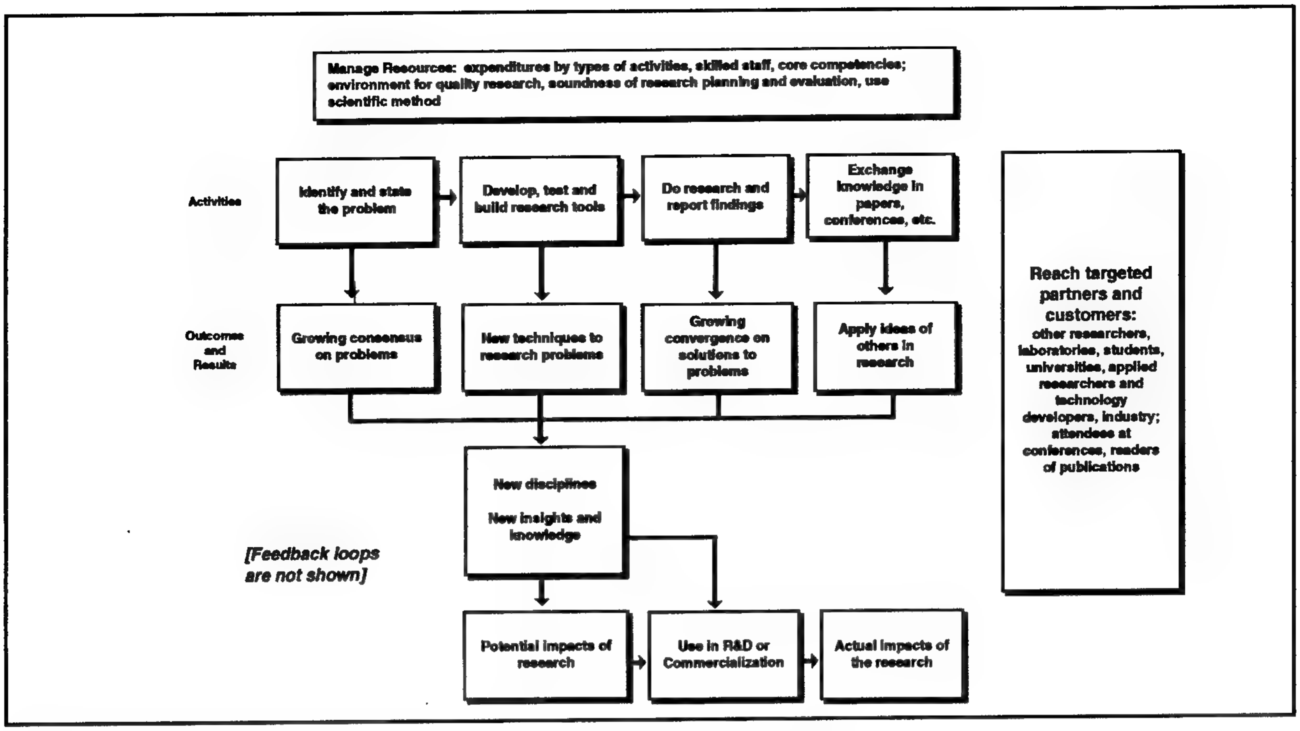

Figure 4. A logical performance story for fundamental research

What will be markers of intermediate progress? Before determining intermediate outcomes, we need to state activities and how these will lead to those intermediate outcomes and then to achieving the objective. There are very basic activities of research which are somewhat sequential, though again certainly not linear: (1) identify the problem, (2) develop the tools necessary to investigate the problem, (3) run experiments to find solutions, and (4) exchange information with other researchers and stakeholders. For each of these activity groups, we can also state generally the sequence of events that might follow, knowing that this is a simplification. For example, identifying the problem requires investigation, and progress can be seen as consensus grows about the statement of the problem. Similarly, experimentation and analysis leads to narrowing the theories and options for solutions before a final solution is found. Notice that all problem identification and experiments are successful when we describe the objective as narrowing the number of possibilities. Development of tools follows a path of design, test, evaluate, revise and utilization that can be measured. Finally, exchanging knowledge involves things that are easy to count such as number of students trained, number of conferences held, number of citations in peer-reviewed journals. The outcomes of the exchange, such as serendipitous discovery, are more difficult to measure.

How are we managing our research resources? The lag between funding and measurable socio-economic impacts is so long with so many confounding factors, it is particularly important to demonstrate good management practices. Quality research means using the scientific method. It also means choosing the right problems to work on, that is, research must be relevant and useful to the customer who is funding the research, if not others. Performance measures also can reflect the historical perspective on research progress, the significance of the problem, and the strategies and skills being applied to solve it.

Who are the people involved? Diffusion of research and support of partners are acknowledged as essential for progress and as evidence of quality and significance. The skill of the researchers, integration along the research continuum, multidisciplinary teams, and collaboration are also considered key factors for excellent, relevant research. Thus, researchers need to demonstrate they are working with key partners who also put resources to the task. Similarly, wherever possible they must define targeted customers for the new knowledge and demonstrate that they are reaching those customers, at least to verify the need for the potential outcome. Mission-directed research can discuss potential impacts if a problem is solved and demonstrate that there are target customers who care that it gets solved. Non-mission research can discuss past impacts on customer groups. Finally, there are stakeholders who are impacted by our expenditures if not our actions. It is often helpful to know where the money is being spent, particularly for the large research facilities.

Step 2. Organizing the performance spectrum in a logical fashion.

Programmatic activities, management of resources, and the people reached by the organization, all together, will lead to the expected results of new insights, new

paradigms, new disciplines, and to potential applications in other research or technology development and deployment program. The particular new insights, disciplines, and potential impacts of any research program can be described in non-technical language. Figure 4 captures this logical performance story in a diagram, from activities to outcomes, with key management and customer issues highlighted in side boxes.

Step 3. Choosing balanced measures for a fundamental research program.

Now the balanced scorecard approach is used to choose measures that tell the performance story for a fundamental research organization. The questions an organization might ask to determine the critical few measures are based on the performance spectrum outlined above and displayed in the logic chart. Possible measures, balanced across the performance spectrum, are listed after the questions. Possible sources for the performance data are also proposed.

Resources Management: Questions to address about management of resources include: (1) What resources do we have, and for what activities are these expenditures being used? (2) What is the problem to be solved and how are projects chosen? (3) Where are you in the process, that is, what are milestones? Data for these measures may come from financial reports, annual progress reports, program managers, peer review, or customer evaluations. Measures that could be used to answer these questions include the following:

- Annual budget by type of activity ($ on facility, $ on strategy/research area)
- Cost share of partners
- Expertise of staff and collaborators
- Indicators of good management of facilities
- Indicators that scientific method is followed
- Indicators that an environment for excellent research is provided
- Milestones and outputs, logically linked to outcomes (counts such as reports, experiments).

Right People Reached: When trying to demonstrate that the right people are being reached by the organization, the questions are: (1) Who is interested in the problem and why? (2) Who are you working with and why is that significant? (3) How many and what types of people are aware of your research? This data could be collected by the program staff, and also asked during customer evaluations. Measures that could be used to answer these questions address the number of targeted populations reached, for example:

- Collaborate with __% of universities doing research in this area
- Published results in journal reaching ___ in areas of __, __, and ___
- Provided training to ___ graduate students.

Research Progress and Results: Questions asked about progress and results include: (1) How do you know you are doing quality work? making progress? (2) Has there been paradigm shift, new discipline, narrowing of areas of solution? (3) Have new tools or methods for research developed? and (4) What are the potential (and actual) outcomes? Peer or expert panel review will be the source of data for most of these measures, with supporting data from program and customer data. Measures that could be used to answer these questions include:

- Indicators of growing awareness of problem or consensus on problem definition
- Tools designed, tested, developed to help address research
- Indicators of paradigm shift or new discipline, convergence on a solution
- Indicators of new knowledge formation
- Indicators of quality science, relevance of research
- Customer satisfaction
- Indicators of potential application and impact
- Use in R&D or Technology development
- Actual impact (if applicable). .

Cost-effectiveness is a thorny issue for fundamental research. Even if programs collected the measures of resource management and results mentioned above, the ratios of input to outputs that follow may not correctly reflect the value of the research or be useful for program improvement. Some examples of possible cost effectiveness measures are the following:

- Dollar value of past impacts compared to dollars of current year funding
- Dollar value of potential impact (weighted by likelihood of success) compared to dollars of current year funding
- Number of persons reached with research findings compared to current or cumulative dollars of funding.

Conclusions and Next Steps

These performance planning tools and the resulting balanced set of performance measures have been well received as we worked with program managers and briefed interested groups. Comments from the EE technology development and deployment programs indicate increasing comfort with the balanced set of metrics emerging in that arena. The generic performance story for fundamental research was presented to the federal Interagency Research Roundtable, and the intermediate outcomes were thought to be particularly useful. As programs personalize the stories and the measures that result, the generic performance stories can be improved. The use of performance data in strategic and operational planning will increase as the measurement improves.

Particularly in the fundamental research area, definition of indicators as proxies for measures, the units of measurement, and measurement methods need to be improved in order to evaluate areas that formerly have not been studied or have not been studied from a holistic, balanced perspective. Our current research is focused on determining measures for the environment for research, thus expanding the resources management question posed in this paper. How can an organization assess its environment for hiring and retaining the best people and facilities, for having an effective infrastructure, for having mechanisms to pursue and validate new knowledge, and for the program planning and external interactions that keep research relevant to science and society? We find this an exciting challenge.

References

Corbeil, Ronald. "Logic on Logic Charts." *Evaluation Newsletter*. Ottawa: Office of the Comptroller General of Canada, September 1986.

Corbeil, Ronald C. "Evaluation Assessment: Case Study of Planning an Evaluation." *Action-Oriented Evaluation in Organizations: Canadian Practices*. Wall & Emerson, Inc.: Toronto, 1992, pp. 107-128.

Hauser, John R. *Metrics to Value R&D: An Annotated Bibliography*. Cambridge, MA: Massachusetts Institute of Technology, 1996.

Jordan, Gretchen B., and Darrell Beschen. "Planning For Evaluation of the U.S. Department of Energy's Energy Partnership/Climate Change Programs." *1995 International Energy Program Evaluation Conference*. Chicago, IL, August 1995, pp. 22-25.

Kaplan, Robert S., and David P. Norton. "Using the Balanced Scorecard as a Strategic Management System." *Harvard Business Review*, January-February 1996.

Montague, Steve. *Performance Framework Approach to Public Management*. Discussion Draft, 1993.

Montague, Steve. *The Three R's of Performance-Based Management and Balanced Performance Scorecard*. 1996. Available from: http://www.globalx.net/pmn/scorecard/index.html: INTERNET.

Tipping, James W., Eugene Zeffren, and Alan R. Fusfeld. "Assessing the Value of Your Technology." *Research Technology Management*, 1995.

U.S. Department of Energy. *Guidelines for Performance Measurement*. DOE G 120-1-5, June 1996.

U.S. Office of Science and Technology Policy. Committee on Fundamental Science, National Science and Technology Council. *Assessing Fundamental Science*. July 1996.

Wholey, Joseph S. *Planning Useful Evaluation: Evaluability Assessment*. Sage: Beverly Hills, CA, 1980.

Author Biographies

The authors are in the Energy Policy and Planning Department of Sandia National Laboratories, one of the U.S. Department of Energy's defense research laboratories. **Gretchen Jordan** has a Ph.D. in Economics, has experience working for Congress and as Chairman of a Business Administration department, and is co-chairman of the American Evaluation Association Topical Interest Group on Research, Technology and Development Evaluation. **John Mortensen** has a Ph.D. in Public Administration from Syracuse University with emphasis in Science and Technology Policy and Development Administration. Before coming to Sandia he worked with city and state governments. Drs. Mortensen and Jordan work with DOE Headquarters staff in the Energy Efficiency and Renewable Energy Office to establish a corporate program evaluation and performance measurement system.

Measuring Performance at the Army Research Laboratory: The Performance Evaluation Construct

Edward A. Brown

Special Projects Office, U.S. Army Research Laboratory
Adelphi, MD 20783-1197

Abstract

The Army Research Laboratory (ARL) is the only R&D organization to be designated a Pilot Project for Performance Measurement under the Government Performance and Results Act of 1993 (GPRA). As such, managers at ARL were required to develop a system for measuring progress towards meeting strategic and annual performance goals, as specified in the required planning documents. The measurement system developed at ARL to comply with the GPRA, which we call the Performance Evaluation Construct, is presented herein.

Introduction

The Army Research Laboratory (ARL) is one of the approximately 70 voluntary pilot projects under the Government Performance and Results Act of 1993 (GPRA), having been so designated by the Office of Management and Budget (OMB) on 6 July 1994. With this designation came the responsibility to develop and publish strategic plans, annual performance plans, and annual performance reports, as well as to design a performance measuring methodology to support this planning and reporting process. All of these planning tools were to be subjected to the scrutiny of OMB, the General Accounting Office (GAO), all levels of headquarters management superior to ARL in the Army and the Defense Department and, ultimately, Congress.

This begs the question of why we would willingly submit ourselves to such "visibility." The answer is that we were already doing all of the tasks that would be required of us, since it simply represented good business practice. In addition, since GPRA is a law that will be implemented government-wide within a few years, we believed that someone had to speak for the R&D community. Quite frankly, we were afraid that if the community didn't come up with ways to plan and measure that were appropriate for the very unique characteristics of research, people who didn't appreciate these subtleties might impose a totally inimical system upon us.

ARL was activated in October 1992 as the consolidation of seven formerly independent (or "corporate") Army laboratories. Our first director took the position that, since we were building an entirely new organization, we should use the opportunity for innovation. Rather than continue operating in the ways the former laboratories managed themselves, ARL should be run more like a private sector enterprise. Specifically, he directed the design of a business planning process that could be used for guiding the laboratory towards both long- and short-term goals, and a performance measurement process that could demonstrate the progress made towards those goals. It is that process that has evolved into the ARL Performance Evaluation Construct—a rational, logical, semi-quantitative methodology that allows the director to offer an assessment of the health of ARL's technical programs and functional operation.

A Discussion of Metrics

When one talks about measuring an organization's performance, the word "metrics" inevitably becomes the focus of the discussion. What is it that makes a good metric and why does research present such a difficult challenge for the application of metrics? First, any metric must have three characteristics: it must be meaningful, it must have an "appropriate" timeframe, and it must have some goal or some definition of what is "good." For instance, in a production or assembly line environment, a meaningful metric could be some dimensional tolerance (plus or minus some number of mils) on the object being manufactured. The timeframe of the metric could be on the order of minutes, if not seconds, because if the dimension started to move out of tolerance, some immediate adjustment could be made based on a measurement of the dimension, and the production line could continue. The definition of "good" would, obviously, be the ideal dimensional size.

Contrast that scenario with a research laboratory where managers have struggled for years, even decades, to define useful metrics. Consider, as an example, the number of patents obtained, which is a frequently suggested metric. Is it meaningful? As a measure of activity, certainly. As a measure of quality of work, or even more importantly, of the impact that the research will eventually have, hardly! Also, the timeframe for obtaining a patent is on the order of three to five years. Should one want to use patents as a management tool, by the time that a fall-off in patents was observed and corrective action was taken and had time to show an effect, half a decade or more could well pass.

Journal of Technology Transfer Vol. 22 (2): 21–26.

Finally, what is the "right number" of patents for an organization to produce per year? To even attempt to answer such a question is foolish.

Why is it so difficult to measure research, a field in which measurement is at the very core? There are a number of reasons, most of them quite obvious. First, the likely outcomes of research cannot often be quantified in advance. The results are often more serendipitous than predictable, and there are usually inputs from many sources which combine with the outputs of the research program that result in some eventual outcome. Also, the high percentage of negative findings, while considered a fundamental part of research, is not a positive addition to an organization's marquee. Second, the knowledge gained is not often of immediate utility; rather, there is often a very long time lag—often several decades—between the inputs/outputs and the outcomes. Finally, and most simply, the unknown cannot be measured.

While it is true that the "D" of R&D is somewhat more amenable to measurement than the "R," until work has progressed well into development towards engineering, the task is fraught with pitfalls. The literature is replete with unsuccessful schemes to measure research going back at least forty years. Most put forth one or another countable as an indicator; many combine several of these in some heroic algorithmic manipulation which, more often than not, is not only unenlightening, but usually obscures whatever small meaning that could be derived from the individual countables standing "unprocessed."

Development of the ARL Performance Evaluation Construct

Aware of the historical problems associated with measuring performance in an R&D environment, we first began our task by defining the "conventional wisdom" for assessing a research program. We found the most meaningful and most respected methodology to be the *retrospective review*: "Twenty years ago we did this piece of research and look what fruits it has yielded today," usually followed by, "So, therefore, Mr. Sponsor, if we keep operating the way we did 20 years ago, in 20 more years you'll reap equally impressive benefits. Send money; have faith!" Many researchers hold that this approach has merit. However, in today's economic climate most sponsors, customers, and stakeholders view this approach as absolutely unacceptable, to the point that such a stance could be very detrimental to the career of a research manager.

Another approach that is still meaningful is the *peer review*, a widely accepted method for determining the scientific merit of a piece of work already completed. (Predictive peer reviewing is also a common tool used in the grant selection process; however, this is not the application of interest here.) Peer reviewing has its limitations, since it may convey to stakeholders that the work being done is *good* work, but not that it is *useful* work. Also, even the estimate of how good the work is can be skewed by an imbalance of representation in the peer review group's membership.

Finally, there are *metrics* which are very timely from a reporting sense, but almost always are measurements of input or activity, or sometimes output, but never of outcomes for the reasons stated above.

With this understanding, we began formulating a new approach by asking the question, "What information does the stakeholder really want to know from a performance evaluation system, beyond what the ultimate outcomes and impacts of the research will be?" We have accepted the fact that outcomes and impacts are, essentially, impossible to determine so far in advance, and have moved beyond attempting to deal with them. However, there are three things our stakeholders want to know, or at least are willing to settle for:

- *Is the work relevant?* That is, does anyone care about what we are doing? Is there an aim or goal, no matter how distant, that our sponsor can relate to?
- *Is the program productive?* That is, are we moving toward a goal, or at least delivering a product (in some form) to our customer in a timely fashion?
- *Is the work of the highest quality?* That is, can we back up our claim to be a world-class research organization doing world-class work?

We used the tools available to us—peer review and metrics—to answer these questions. Furthermore, we realized that if we could define our customers or stakeholders, evaluation or feedback from them would be another useful tool. We felt that if peer review was independent and of sufficient stature it would answer the question concerning quality, and that customer evaluation would certainly speak to the issues of relevance and productivity. After all, the person paying the bill for the research was a customer, and if the customer had no requirement for what was being done, or wasn't getting anything from the program, the paychecks would, in all likelihood, soon stop. Metrics could be considered an adjunct to the two other methods. The metrics tool has become quite useful to ARL, but in a somewhat unexpected way, which will be described later in this paper. Figure 1 shows the development of the Construct in the form of a matrix.

Implementation of the Construct

Peer Review

The current director of ARL was formerly the director of the National Institute of Standards and Technology (NIST). He brought with him to ARL the concept of the NIST Board on Technical Assessment, a peer-review group established and administered by the National Research Council (NRC) of the National Academies of Sciences and

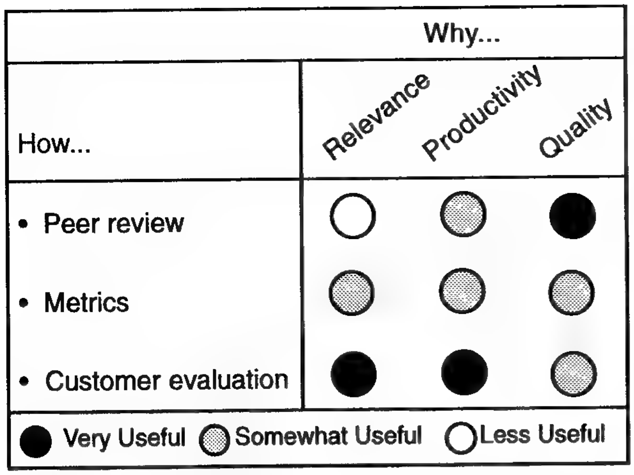

Figure 1. The ARL performance evaluation construct

Engineering. Over the past year we have contracted with the NRC to establish a similar group called the ARL Technical Assessment Board (TAB). The Board itself consists of 15 world-renowned scientists and engineers. Under the Board are six panels of seven to ten individuals with equally fine reputations. These panels, one for each of ARL's major business areas, have the responsibility to pay an annual visit to their respective ARL directorates, spending several days being briefed, touring the facilities, and meeting with the technical staff. Based on their findings, panel members write a chapter of an annual report, which is assembled by the Board and published by the NRC as a public document.

The three purposes of the TAB are to review the scientific and technical quality of ARL's program, to make an assessment on the state of ARL's facilities and equipment, and to appraise the preparedness of the technical staff. TAB members can make judgments about how our program compares with similar programs elsewhere, and suggest improvements. They can also undertake special studies for the director. However, they are specifically enjoined from offering opinions on programmatic content and issues. The director feels he can get advice on that subject at any street corner in Washington. What the TAB supplies that he *can't* get elsewhere is a technical appraisal of the scientific merit of the ARL program.

The TAB hss now completed its first year of operation. The panels visited the Lab during the summer of 1996, and delivered their draft chapters to the Board, which edited and compiled them into a complete report. The NRC subjected the report to its formal editorial and referee process. The Board released the document to ARL's Director during a meeting, at which time their assessments and suggestions for improvement were briefed. The report was then published as an open-literature document and forwarded to the senior leadership of the Army and the Defense Department, as well as to other Government and private-sector leaders. The timing of the report's release was scheduled to just precede the annual ARL strategic planning meeting held by the Director for his senior management, so that its findings could be used as inputs to the formulation of adjustments to the long range strategic plan.

ARL has a contract with the NRC based on a scope of work that delineates the above process. The NRC supplies a full-time staff director and several part-time assistants to administer the TAB, and selects the members of the Board and the panels. The members receive no salary, but their expenses are paid out of the funds received by the NRC from our contract. Because of the size and diversity of the ARL program, the TAB can only assess about one third of the program each year, so it will take three years before the total program is completely covered once.

The independence of the NRC's operation, the international stature of the members of the Board and the panels, and the imprimatur of the National Academies brings a validity to the process that our stakeholders accept as a valid response to the quality question of the Construct.

Customer Evaluation

Stakeholders of the research enterprise. Before discussing ARL's approach to customer evaluation, it is necessary to identify our customers and what they expect from us. Dr. Edward B. Roberts of the M.I.T. Sloan School of Management has developed a model of the stakeholders in the research department of a firm in the private sector. According to Roberts, there are three groups of these stakeholders. The most readily apparent group is the development and manufacturing departments, which are directly dependent on the results of research for production of their new products. The second group, external to the firm, is the end item user, the company's customer for finished products. Although the researchers usually don't interact directly with the user, they must keep in mind that the end product of their work will eventually lead there. Finally, the third group is again within the firm, and was not as immediately obvious to us. It is the senior leadership of the firm—the CEO, the Chief Technical Officer, and others at that level. These individuals are also stakeholders in the research enterprise because they are critically dependent upon the achievements in their laboratories for strategic planning purposes.

These three groups translate well into our Army corporate laboratory environment. Within the Army's acquisition community, the first group translates into the Research, Development and Engineering Centers (RDECs), as well as the program managers and program executive officers (PM/PEO), all of whom are concerned with the development and procurement of materiel for the soldier (e.g., helicopters, armored vehicles, communications, weapons, etc.). The second group, the end-item user, becomes the soldier in the field, as represented to us by the Army's Training and Doctrine Command (TRADOC). It is TRADOC's responsibility to define how the soldier will fight, how he will be trained, what doctrine he will use, and

what materiel he will need. There is a strong "push-pull" tension between the materiel under development and the determination of future doctrine, each one influencing the other. The third group is the Army's senior leadership, the Chief of Staff, the Vice Chief of Staff, and others. These individuals rely, at least in part, on what ARL is investigating now that will reach the field in 10 or 20 years, and what effect those developments will have on the force structure of the future Army. For instance, if ARL can successfully develop the technologies that will enable a 2-person, 40-ton tank to replace the current 4-person, 70-ton M1A1, the impact on the future force structure and doctrine could be enormous.

Feedback from the first group of stakeholders: the RDECs and the PMs/PEOs. Our methodology for obtaining feedback from the RDECs and the PMs/PEOs is relatively straightforward, since we deliver something tangible to these customers—a report, or a device, or a model, or a program. Although there are several subgroups within this class of stakeholder that require somewhat different handling, in all cases there is some written documentation or scope of work that defines what is expected of us, over what time period, and for what cost. The signatories to this documentation are usually at the first- or second-line supervisory level, both at ARL and the customer. Therefore, with a transmitter and a receiver so identified and the product well-defined, the tool of choice is, quite simply, a survey. We designed a simple five-question instrument, with additional room for comments. Answers are given on a one-to-five scale, five being the highest score. Questions are asked to determine if the product delivered by ARL was what was agreed upon, if it was timely, if it worked, etc. The director's policy is that every customer is to be polled annually, and that any individual question marked by the customer as a 1 or 2, or any comment with a negative tone, will be responded to personally by the head of the directorate in which the work was done within five working days. This requirement is included in the performance standards of each of the directorate heads. The director reviews all the unfavorable responses personally.

We have used this survey process for three years, with approximately 400 surveys sent out each year. The response rate is about 60 percent, and the overall score for ARL has been climbing from an initial 3.9 to a current value of 4.3.

This survey approach is, obviously, more appropriate for the applied research conducted by ARL. The more fundamental work is, not surprisingly, somewhat more problematic. In many cases, the customer for this work is the director himself, or sometimes another group within the laboratory that will apply the basic findings to another project. In cases such as these, the director is able to provide feedback in an immediate, sometimes dramatic, fashion without having to rely on survey results.

Feedback from the user and senior leadership. Since ARL does not deliver any tangible object to the Chief of Staff of the Army or others at that level, the use of a survey is obviously inappropriate. Nevertheless, it is critical that we understand how our work is perceived at that level. If our work is not viewed as adding value to the Army by these important stakeholders, our very existence as an organization is quite literally in jeopardy. In order to ensure that ARL remains closely coupled to the Army's vision and responsive to the senior leadership, and especially to ensure that we continue to be viewed as such, we established a Stakeholders' Advisory Board (SAB). The SAB is chaired by the Commanding General (four-star level) of our parent organization, the Army Materiel Command, and comprises ten members of the Army's senior leadership at the three-star (or equivalent civilian grade) level. This group meets once a year to review ARL's program from a broad strategic level. They do not involve themselves with the details of the technical program; rather, they are concerned with whether the total thrust of the program is responsive to the needs of the Army. The group discusses such issues as whether the overall program is in balance from several dimensions (e.g., mission-versus-customer funding, in-house work versus contractual programs, near-term emphasis versus far-term emphasis, etc.), and deals with high-level relationships between ARL and other Army organizations. The SAB can give guidance on broad areas in which they believe more emphasis will be needed in the future, such as the fact that there will be an increased need for enhanced communications and information processing on the future battlefield, which requires an increased emphasis on digital technologies. The SAB has met once; its second meeting was in July 1997.

Metrics

The third pillar of the Construct is metrics. As previously explained, we do not believe that there are any valid metrics that can adequately or usefully describe the technical performance of a research organization. However, there are two important roles that metrics can play. The first role is as an indicator of the operational or functional health of the organization. There are certain practical, as well as legal, constraints under which a Government organization must function. Also, the environment or climate can be more or less conducive to promoting excellence in research. Metrics can be useful indicators of these factors. Variables such as overhead rate, average age of the workforce, cycle time for small-purchase procurement, and a myriad of others are examples of factors that can be monitored through the use of metrics. Presently ARL tracks 54 metrics of this sort, but most of the resulting data are never reported to the director or to any others outside of the functional offices responsible for collecting the data. As long as these metrics stay within the usual, predetermined bounds, no problem exists. However, if a metric deviates too far from the norm, that fact can be brought to management's attention for corrective action. It is in this

	Relevance	Productivity	Quality	Functional Health
• Peer review			TAB	
• Metrics				Metrics in Performance Standards
• Customer evaluation RDEC/PMs Users/Leadership	Surveys SAB	Surveys SAB		

Figure 2. Implementation of the ARL performance evaluation construct

spirit that we count the old standbys—papers, patents, presentations, and so forth. We do not use these metrics, or any others, to determine an annual "score" for ARL. Whether or not we are awarded 100 patents or 110 patents in a year is not critical *per se*. However, not receiving *any* patents would be an indicator that something is awry and warrants the director's intervention.

The other use that we found for metrics is as a tool to help the director "steer" the laboratory in directions that he wants it to go. For instance, it is his belief, based on his experience with other world-class research organizations, that 40 percent of the technical staff should be educated to the doctoral level. Upon his arrival at ARL, the director found that only 22 percent of the technical staff had earned doctorates (due, in part, to the historical fact that the constituent organizations that formed ARL were more engineering-oriented). Therefore, he set a long-term goal of 40 percent, with intermediate goals leading up to that figure. He then added these goals to the performance standards of his senior managers, as an impetus for them to hire (to the extent allowable) or train individuals to the doctoral level. Another example of the director's use of metrics is that, again based on his personal experience, he believed that every scientist and engineer should publish at least one journal article or technical report per year. We have a little over 1500 members of the technical staff, so he set an ARL-wide goal of 1500 publications (divided between papers and reports) and apportioned these goals to the various directorates according to the nature of their programs (i.e., the directorates with a higher proportion of basic research work were given a higher goal for papers and a lower goal for reports, with the opposite situation for the more applied directorates).

Of the 54 metrics being tracked, 15 are currently on the director's "short list," and goals for their completion are included in the standards for the senior managers. These metrics are collected solely at the pleasure of the director, and may change from year to year. Trend data is only important to the extent that the director deems it so for his own purposes. Goals are set by him based on his own experience and the results of benchmarking studies that we have done. For instance, the Naval Research Laboratory, an organization comparable to ARL and one of world-class reputation, which we would like to emulate, has a technical staff with twice the percentage of Ph.D.s as our laboratory. This information provides us with a benchmark for that metric.

Summary

The current implementation of the ARL Performance Evaluation Construct is shown in Figure 2. It is, we believe, as close as we can get to being able to report on ARL's status and performance, recognizing that truly reporting outcomes and impacts of the research program is not feasible. The three pillars of the Construct come together, not in a numerical score derived through some arcane algorithmic process, but rather in the director's head as he integrates the results of the peer review by the TAB, the customer feedback from both the surveys and the SAB, and the metrics. The director can gather all the details, process them, and make a determination as to whether or not the laboratory is operating satisfactorily. He is then able to present his findings, with the available evidence, to any audience in any format required, tailoring the presentation as appropriate. The director is then able to respond informatively to the questions of relevance, productivity, and quality. What he doesn't do is just give this year's score!

Author Biography

Dr. Brown is currently the Director for Special Projects at the U. S. Army Research Laboratory (ARL). He is the project manager for its participation as a pilot project under the Government Performance and Results Act of 1993, and as a National Reinvention Laboratory under the National Performance Review. He is responsible for much of ARL's groundbreaking work in performance measurement and business planning as it applies to R&D organizations.

FY97 ARL Performance Metrics
Grouped by Vision elements

Preeminent in key areas of science.

- Deliverables
  - Top Tasks (% met) √
  - STO's (% met) √
- Documentation (leaving tracks)
  - # Refereed paper/proceedings √
  - # Internal ARL Technical Reports √
    - # Formal Surv./Lethal. Reports (SLAD)
    - # of completed software packages (ASHPC/IST)
  - # Chapters/books written
  - Patents
    - # total
    - # Invention Disclosures
- Facilities/Equipment
  - $ Value capital equipment purchased in FY
  - $M invested in facilities in FY
  - Replacement rate of facilities

Staff widely recognized as outstanding.

- Profile
  - %PhDs (S&Es) √
  - # Technicians per S&E
- Training
  - % Emp. with 40+ hrs training
  - # Emp. on long term trng
  - PhD candidates
- Esteem Factors
  - # Significant awards
  - # Invited Presentations
  - # Prestigious Posts

Miscellaneous

- Financial
  - Obligation
  - Disbursement
  - IH/OGA/Contract $s
  - Indirect Overhead ($M) √
  - G&A (% total revenue)
- Personnel Statistics
  - Glidepath √
  - Avg age (S&Es; total)
  - Avg grade (S&E; total)
  - Avg sick leave use (S&Es; total)
  - Turnover rate (S&Es; total)
- Procurement
  - Avg small purchase cycle time
  - % of HEI (HBCU/MI) contract $s
  - ALT/PALT

Seen by Army users as essential to their mission.

- Technology transitions
  - # Significant technology transitions
- Ratings from customer surveys
  - TPAs √
  - Reimbursable customers √
  - Users
  - Senior Leadership
- Financial
  - Reimbursable Customer Orders ($M) √
- Greening the Workforce
  - # Officers
  - # Enlisted
  - %Emp. completing "Greening Course"
  - # Employees completing FAST, Jr. training
  - # FAST advisers

Intellectual crossroads for the technical community.

- Guest Researchers out of ARL
  - Total # √
  - Total myear equivalents √
  - Avg length of stay
  - # staying 3+ months
- Advisers
  - # NRC Approved Advisers
- Guest Researchers into ARL
  - Total # √
    - # Post-docs √
    - # from HBCU/MI
  - Total myear equivalents √
  - Avg length of stay
  - # staying 3+ months
- Cooperative R&D
  - # new CRDAs
  - # new PLAs
  - Income from CRDAs/PLAs
  - # TPOs/#ATPOs (Int'l)

Key

■	4	vision elements + misc.
➲	17	subcategories
√	15	metrics included in Director's Perf. Stand.
	54	metrics total

Developing and Transferring Technology in State S&T Programs: Assessing Performance*

Julia E. Melkers

Department of Public Administration and Urban Studies, School of Policy Studies, Georgia State University
Atlanta, GA 30302

Susan E. Cozzens

The National Science Foundation and Department of Science and Technology Studies, Rensselaer Polytechnic Institute
Troy, NY 12180

Abstract

By 1996, all states had established a program focusing on the development of technology and technology-based economic development. As more agencies move to performance-based management, state S&T programs are increasingly under pressure to report outcome and output data for their programmatic activities. This paper presents findings on the extent and use of performance measurement and evaluation efforts in state science and technology programs. The 1995-96 study was based on a series of eight case studies and a mail survey of science and technology-based programs in all fifty states. The findings show that three groups of measures emerged as most important to state science and technology programs: employment-related data, leveraged or matching fund data, and anecdotal evidence. State programs are especially pressured to report short-term outcomes, yet show economic benefits. Many state program managers find value in performance data—the research shows that the primary reason that many states assess their performance is the value of performance information as a management tool.

Technology Transfer in the States

Technology transfer initiatives in the states have mirrored federal efforts, implementing programs that attempt to stimulate technology-based economic development (Eisinger 1988). The development and transfer of technology at the state level is achieved through a variety of outreach mechanisms, including (Coburn 1995, p. 17):

- **technology development**: research and applications for new or enhanced industrial products/processes;
- **industrial problem solving**: identifying and resolving company-level industrial needs through technology and best-practice applications;
- **technology-financing**: public capital or help in gaining access to private capital;
- **start-up assistance**: aid to new small technology-based business; *and*
- **teaming**: help in forming strategic partnerships and alliances.

The rationale for technology transfer programs at the state level is that state programs may facilitate the transmission and diffusion of new technologies from the lab or entrepreneur to the private sector. In turn, these technologies can become the impetus for new business creation, the introduction of new product lines to selected firms, or the revitalization of mature industries (MN Department of Trade 1988.) Generally, technology transfer is not explicitly defined as part of the state program mission. Instead, programs are defined as supporting technological development and economic development. In 1995, 49 states have sponsored cooperative technology programs (Coburn 1995). These programs emphasize private sector and university-led efforts, through support from state governments, with goals of the development and diffusion of technologies.

In 1988, 43 states had at least one program encouraging technological innovation (MN Department of Trade 1988). Of these, 26 of these states had a technology transfer mission as part of their technology programs. By 1996, all states had established a program focusing on the development of technology and technology-based economic development. Many states have more than one program and all states have established some cooperative venture, drawing from private sector and university-based assistance. State programs tend to draw primarily from expertise within the

*Any opinions, findings, conclusions, or recommendations here are those of the authors and do not represent the views of the National Science Foundation. This research was supported under NSF Grant No. SBR-9422433 from the Program of Research on Science and Technology and through the second author's NSF Independent Research Plan. The research benefited greatly from the contributions of two tireless research assistants, Arlene Allen and Michael Turton.

state, including industry, university researchers or federal laboratories located in the state (Coburn 1995). Some argue that state leaders are acting less on the basis of economic theory than on the political need to appear to be doing something (Feller 1988; Anton 1989). Our research showed that this political pressure has important implications for the development of performance measures for these programs.

David Osborne (1990) describes these programs as a mixture of successes and failures, due mostly to their experimental nature. Assessing the performance of these programs, however, has been sporadic. As states are increasingly involved in cooperative programs to support the development and diffusion of technology, questions of their performance arise. Identifying appropriate performance measures for state technology-based programs is complicated by the diverse and multiple missions of these programs—to develop technology, create local and regional economic development, and support state universities. Irwin Feller notes that state evaluation efforts appear to use a mix of process and outcome indicators (Feller 1988). There has previously been little systematic information about the actual performance measurement processes, measures, and activities in these state-level programs. We present research findings on the extent and use of performance measurement and evaluation efforts in state science and technology programs. The 1995-96 study was based on a series of eight case studies and a mail survey of science and technology-based programs in all 50 states.

Assessing State and Federal Level Technology Transfer Success

Simply getting technology into the hands of entrepreneurs will not necessarily lead to a successful product or economic benefits. Even if a successful product is developed, in technical fields, firms need to continually refine the process by which they manufacture their products in order to be competitive (Melkers et al. 1993). While government programs address the initial transfer of technology that leads to new products, they do not always address the process improvements that are needed by firms in order for them to maintain a competitive status (Melkers et al. 1993). Some may argue that this is not the task of government; yet, clients of some state-based business assistance programs express the need for assistance in basic business management skills, including development of marketing and business plans and cash flow management (Melkers 1997).

The relationship between technology transfer and economic development is not straightforward. Due to their very nature, the development of technologies and technology transfer should not be viewed as activities that will necessarily produce immediate economic development benefits. A multitude of factors impact whether or not a technology will result in economic growth, many of which lay beyond the control of policy-makers or program staff. Simply nurturing start-ups does not insure that diffusion will occur. Firms need technical capacity, suitable organizational structures and processes amenable to innovative behavior, and a way to analyze and understand market signals (Melkers et al. 1995). In terms of technology transfer, government can and should play an important role in lowering the barriers to cooperative R&D and providing the infrastructure and incentives by which technological progress and subsequent economic impacts may be achieved. Given this, appropriate performance measures will reflect the ability of state programs to facilitate development and transfer of technologies, and will not focus only on long-term outcomes. A constant struggle in the evaluation of technological activities is the identification of appropriate and reasonable indicators. The same is true for economic development programs (Hatry et al. 1990). It is especially problematic to develop appropriate performance measures that represent not only the long term outcomes of technology-based economic development programs, but also take into account more immediate economic and scientific outputs. The problem arises in the tension between the need to report outcomes and outputs and the long term nature of both technological and economic development.

The reasons that state S&T programs establish performance measures and a systematic approach to collecting and maintaining performance measures vary. First, states may be required to produce performance measures and data as part of legislative guidelines. Recent work shows that in 1996, 32 states had legislation requiring performance measures of state agencies (Melkers and Willoughby 1996). Further, ten states had other performance-based initiatives, either in the form of Executive Order or initiative from a central budget agency in the state. Second, another and often more important reason that state S&T programs maintain performance data is due to internal motivation to do so. Management can develop the culture of an organization to be receptive to performance measures and outcome-based management. These two rationales illustrate the difference between a program justification approach where protection of the program is a key motivation, versus a program improvement approach, where performance information is viewed as useful to program management.

Research Design

Findings of this research are based on case study and mail survey data of managers of state science and technology-based programs. The research was conducted in two major phases—a series of intensive case studies and a mail survey. The case studies were intended to gather rich qualitative data on the environment for performance measurement in the state S&T programs, including information on measures, measurement processes, and uses of performance measures and evaluation data. For the case studies, in-person interviews were conducted with indi-

Table 1. **Case studies**

State	Program
Arkansas	Arkansas Science and Technology Authority (ASTA)
Indiana	Indiana Business Modernization & Technology
Kansas	Kansas Technology Enterprise Corporation (KTEC)
Minnesota	Minnesota Technology Inc.
Oklahoma	Oklahoma Center for the Advancement of Science and Technology (ASTA)
Pennsylania	Ben Franklin Partnership
Texas	Advanced Research Program/Advanced Technology Program
Utah	Utah Technology Finance Corporation Centers of Excellence Program

viduals including program managers and staff, members of the programs' boards of directors, legislative liaisons and state budget analysts. The case analysis also involved review of program documents, performance-based databases, and related materials. Table 1 lists the eight case studies conducted in this research.

The purpose of the mail survey (conducted in 1996) was to gather comprehensive information on performance measurement activities in state S&T programs and on perspectives and motivation of program mangers in regard to performance measurement activities. Surveys were sent to program directors of 75 science and technology-based programs in 50 states. Surveys were received from 44 programs (53% response rate), representing 38 states. The survey was designed to gather data on existing or planned evaluation and performance measurement activities. However, the survey also contained a series of questions to gather program director perspectives on the value of performance measurement. Information on perspectives on evaluation in a more general sense was especially important for programs that had a lower level of performance measurement activity.

The 44 state S&T programs that responded to our survey represent a broad range of programs in terms of size, age, and type. Programs are either independent state agencies or programs within a larger agency, such as a Department of Commerce or Department of Economic Development. For the most part, the programs represent a mix of activities including providing research grants, managerial and technical assistance, and coordination of seed and venture capital. Half of the respondents' organizations were created between 1981 and 1989.

The Extent of Performance Measurement and Evaluation in the States

Most state S&T programs collect and maintain some level of performance data. The majority of state S&T programs that responded to the survey either collect performance measurement data on an ongoing basis or have had a one-time evaluation of their program. When asked how they find out and report on how well their activities are doing, most survey respondents indicated they collect performance data. Further, almost half (46%) of respondents collect performance information on a regular ongoing basis. One-third of respondents indicated that they both had had a large one-time evaluation and did regular ongoing collection of performance data.

An important challenge for state S&T programs is the balance between pressure to report short term benefits of program activities and the long-term nature of many of these same activities. Further, many state programs are still new to systematic performance measurement data collection and legislative requirements for performance measurement. The most common measures used in state S&T programs represent a mix of process, demand, and outcome measures. *Outcome measures* are the performance measures that not only receive the most attention, but are also those that are generally most directly linked to programmatic mission and goals. For state level S&T programs, common outcome measures include job generation figures, new businesses started, patents awarded, and cost savings. *Output measures* are measures that represent the generally immediate outputs of programmatic activity. Examples of output measures might include the number of clients served, amount of leveraged funds, and business plans completed. *Process measures* are those that illustrate the level of activity involved in implementing the program itself. Examples of process measures might include number of grant applications processed; number of seminars given, and so on. In addition, one program collects "Program Demand Measures." These are a form of process measure, but are directed at client demand for program components and services. As shown in Table 2, the most common measures used in state S&T programs represent a mix of process, demand, and outcome measures. For much

Table 2. Performance measures reported in state science and technology programs

Measures	Reported Collection (n=44)
Number of projects the organization has funded	68.20%
Matching/leveraged funds	68.20%
Jobs created/new jobs	65.90%
Number of organizations assisted	61.40%
Number of requests for assistance	56.80%
Spinoffs/new firms	56.80%
Patent/license application/receipt	54.50%
Jobs retained	54.50%
Customer satisfaction measures	50.00%
New products	50.00%
Average salary of jobs created	43.20%
Increased sales	45.50%
Cost savings/cost avoidance	43.20%
Average salary of jobs retained	36.40%
Number of collaborations	36.40%
Profits	34.10%
Number of publications	34.10%

federal R&D, measures of success gravitate toward measures of scientific and technological, rather than economic, success. Therefore, measures such as patent generation, publications, and citations are considered valid indicators of R&D activities. At the state level, however, the economic development mission of most programs makes these measures problematic. The problem arises in the ability of the policymaking community to interpret these measures while still assessing the economic benefits of the program. Few programs even track measures such as publication and citations.

Overall, three groups of measures emerged as most important to state S&T programs: employment-related data, leveraged or matching fund data, and anecdotal evidence. First, for most state programs, employment-related data are considered to be the single most important indicators of program performance. The quality and the specificity of the employment data, however, vary by program. Generally, programs are pressured to produce data on jobs that have resulted directly from their program. The challenge arises in the view of the primary programmatic goals and mission. From a legislative point of view, employment is often seen as a primary goal. However, from the perspective of the grantees of many of these programs, scientific and technological progress and success is viewed as appropriate performance measurement. With the exception of programs that use an economic multiplier model to estimate job generation and growth, programs rely on self-reported employment data, generally from client firms. This is certainly the easiest and least expensive source for these data. However, this sort of reporting also presents problems of consistency in reporting, interpreting new or saved jobs, and the link of the new job to the funded project. The specificity of job-related data also varies by program. Some programs report overall job generation, whereas others report on average wage. The less tangible measure of "jobs saved" is also used by several programs. These measures are extremely problematic due to their speculative nature. Not all program staff agree that job generation is an appropriate measure of program performance, especially in the short term. Yet short-term results are most relevant to the legislative and political audiences for performance measures. An important characteristic of science and technology-based programs is that the likely and expected outcomes of the program are generally of a long-term nature. However, legislators must make budgetary and other decisions about state program on an annual or biennial basis.

A second type of measure identified as not only appropriate but also extremely important in representing performance was the level of leveraged funds. Many programs have a matching fund provision which requires grantees to obtain additional financial support from an external source. These funds are important to the state because they represent additional dollars brought into the program. From a performance perspective, the level of leveraged funds represents the ability of a project or series of projects to attract additional support. This is, in effect, an intermediate measure of performance since it not only illustrates available resources to contribute to the likelihood of ultimate success, but also serves as evidence of an external positive opinion of the project's likelihood of success.

Finally, the third most often mentioned performance information (in addition to job generation and leveraged/matching funds) was anecdotal evidence. The rationale behind performance measurement activities is to provide consistent, generally quantitative, reliable evidence of program performance. However, program staff in each of the case studies explained that anecdotal evidence, often including newspaper reports and testimonial letters of support, were critical in demonstrating program performance. Legislators and the public were seen as the audiences for the anecdotal information. Another category of measures worth mentioning are measures of customer satisfaction. Many state governments are undertaking a "quality initiative" that often involves demonstrating response to customer needs. This is reflected in the fact that half of the

Table 3. Output and outcome measures in the case study states

Measures	AR	IN	KS	MN	OK	PA	TX	UT
Jobs created/new jobs	X	X	X	X	X	X	X	X
Average salary of jobs created				X		X	X	X
Jobs retained		X	X	X	X	X		
Average salary of jobs retained						X		
Spinoffs/new firms						X	X	X
Patents/licensing	X		X		X		X	X
Matching/leveraged funds	X		X		X	X	X	X
Increased sales		X	X	X	X			
Cost savings/cost avoidance		X	X	X				
New product development	X		X	X				
New products commercialized			X			X		
Number of publications					X		X	
Number of collaborations					X		X	
Increased capital spending		X	X					
Customer satisfaction measures			X	X				

survey respondents reported collecting customer satisfaction measures.

Measures reported through the case studies are consistent with the survey results. There was a distinct trend, both within some programs and across the set of programs, toward indicators of economic development and the success of target business firms (see Table 3). Again, the most common were measures of job generation and leveraged funds. In the interviews, anecdotal evidence was consistently highlighted as critical in accurately demonstrating performance. This was especially important in reporting program outcomes and outputs in specific geographical areas, such as particular legislative districts.

The Motivation to Assess Performance

As more states adopt performance-related legislation, it seems likely that more science and technology-based programs will collect and maintain performance data on their activities. In many states, the adoption of performance-based legislation requiring performance reporting is the impetus for performance measurement activities. However, although almost half of the respondents were in states that have some form of performance-based legislation, only 16.1 percent indicated that the primary reason that they assess the performance of their science and technology-program is legislative requirements. For those states that reported having a performance-based budgeting requirement, none reported it as having "a lot" of influence on their performance measurement activities. The primary reason, cited by half of the respondents, that they assess their performance is the value of performance information as a management tool. The second most important reason was the use of performance information to justify the program to outsiders. Justification of the program to outsiders generally refers to the need to justify program activities and outcomes to the policymaking community, particularly the legislature.

If state programs feel threatened in the budgetary process, they may look for ways to protect themselves. When asked to indicate how much they agreed with the statement "Our organization is safe from budgetary cuts, at least for the time being," most respondents disagreed or strongly disagreed. Justifying program performance to legislators is one way to reduce the likelihood of budgetary cuts. Evident in most of the cases was a tension between the need to justify the program to outsiders and policymakers and the desire to develop performance measures that are useful for program improvement. The need to justify programmatic activities to outsiders takes a particular form: it pressures the organization to meet policymaker demands and report job generation data, the performance indicator most in demand by legislators.

Performance Measurement in the States

In sum, the research shows that performance measurement activities do exist on a broad scale in state S&T programs. The emphasis on particular measures, such as job generation and leveraged funds, are somewhat consistent across state programs. While new jobs and the establishment of new technology-based businesses may be expected in the longer term, short term information needs are in conflict with what can actually be demonstrated. This presents an important challenge for the state programs that will be key to the acceptance and understanding, by stakeholders, of ongoing performance of these programs.

The pressure to report short-term results, with pressure to show economic benefits, is one of the most critical reporting issues facing many state programs. From the research, it is clear that states that have made a stronger effort to *educate the audience* for their performance information are under less pressure to report job generation as the sole evidence of their performance. Those programs that work more closely with legislative staff and educate them about the nature of science and technology and their relationship to the economy are more successful in reporting a broader range of performance data.

The attitude of program managers about performance measurement and evaluation is important to the use of the performance data. Gaining commitment to performance measurement from key players both inside and outside the organization is important. In many states, performance legislation is in place that formalizes this commitment. However, key players within the agencies, departments, and programs are critical to the success of a performance measurement system. This requires participation of state science and technology program staff, clients, and related policymakers. It also requires a balance between short-term and long-term measures of success.

References

Coburn, C. *Partnerships: A Compendium of State and Federal Cooperative Technology Programs*. Columbus, OH: Batelle Press, 1995.

Eisinger, P. *The Rise of the Entrepreneurial State: State and Local Economic Development Policy in the United States*. Madison, WI: University of Wisconsin Press, 1988.

Feller, I. "Evaluating State Advanced Technology Programs." *Evaluation Review* 12, 1988, pp. 232-252.

Hatry, H., M. Fall, T. O. Singer, and E. B. Liner. *A Manual: Monitoring the Outcomes of Economic Development Programs*. Washington, DC: The Urban Institute, 1990.

Melkers, J. E. *Alaska Business Assistance and Loan Programs: An Assessment of Small Business Needs and Experiences*. Final Report to the State of Alaska Department of Commerce and Economic Development, 1997.

Melkers, J. E., D. Bugler, and B. Bozeman. "Technology Transfer and Economic Development." In *Theories of Local Economic Development*, ed. R. Bingham and R. Meir. New York: Sage Publications, 1993.

Melkers, J., and K. G. Willoughby. "The State of the States: Performance-Based Budgeting Legislation." Unpublished paper presented at Annual Meeting of American Evaluation Association, Atlanta, GA, November 8, 1996.

Minnesota Department of Trade and Economic Development, Office of Science and Technology. *State Technology Programs in the United States: 1988*. St. Paul, MN: MN Dept of Trade, 1988.

Osborne, D. "Refining State Technology Programs" in *Issues in Science and Technology*, Summer 1990.

Author Biographies

Susan E. Cozzens is Director of the Office of Policy Support, and Professor of Science and Technology Studies at Rensselaer Polytechnic Institute. She has published widely on research assessment approaches and methods, and has been an invited speaker on this topic in Europe, Asia, and Australia. She has been a consultant to the Institute of Medicine, General Accounting Office, and National Institutes of Health. Her office is currently coordinating the National Science Foundation's response to GPRA.

Julia Melkers is Assistant Professor of Public Administration and Urban Studies in the School of Policy Studies at Georgia State University. Her research interests include science and technology policy, performance measurement and evaluation at the local, state, and federal level, and evaluation utilization. Dr. Melkers' current research focuses on performance measurement in state science and technology programs.

R&D Value Mapping: A New Approach to Case Study-Based Evaluation*

Barry Bozeman
School of Public Policy and the Technology Policy and Assessment Center, Georgia Tech
Atlanta, GA 30332

Gordon Kingsley
School of Public Policy and the Technology Policy and Assessment Center, Georgia Tech
Atlanta, GA 30332

Abstract

This study presents an approach to harnessing the power of case studies for research evaluation called R&D value mapping (RVM). *While this method uses case studies in the traditional manner to provide in-depth insights, it also structures case studies through an analytical framework that yields quantitative data and less subjective "lessons learned." When properly applied, RVM can yield an inventory of outcomes and empirical generalizations regarding the determining variables. A particular advantage of the approach is that it not only provides an indication of the type and amount (though not a single numerical index) of outcome, but also gives insight into the reasons outcomes are achieved. Thus, RVM is useful for policy management strategies seeking to replicate success. The specific steps associated with the RVM method are illustrated through studies that have applied the technique.*

The set of analytical tools for assessing the social and economic impacts of R&D has expanded significantly during the past ten years. Not so long ago, evaluation of R&D impacts and technology development seemed equal parts alchemy and vaguely derived numbers. As a result of methodological developments, the numbers are currently derived with a bit more rigor. While alchemy still holds sway, serious evaluations are much more common.

Despite advances in application of such research evaluation techniques as cost-benefit analysis (Averch 1994), benchmarking (Rush et al. 1995) and bibliometrics (Rao 1996), one set of obviously relevant techniques, case studies, has remained somewhat stunted in its development. Case study approaches to research impact evaluation generally have credibility with policy-makers and officials and are popular among evaluators and policy analysts (Kingsley 1993). But with the conspicuous exception of the methodological advances provided by Robert Yin (1994), case study approaches to research evaluation remain fragmented, piecemeal and difficult to aggregate. Case studies, in research evaluation as elsewhere, seem to "tell us more and more about less and less." Case studies provide richness and depth of understanding but, all too often, one is left to one's own devices in trying to determine "what it all means." While case studies can provide important lessons, the lessons depend as much on the interpretive ability of the reader as the science of the evaluator.

The objective of this paper is to outline advances in a new approach to harnessing the power of case studies for research evaluation, an approach that has promise, if successful, of using case studies in the traditional manner to provide in-depth insights; but, at the same time, it may use case studies in an analytical framework that yields quantitative data and less subjective "lessons learned."

The method, termed *R&D value mapping* yields an inventory of benefits and empirical generalizations of the determinants of those benefits and has been applied in several studies (Bozeman et al. 1992; Bozeman and Roessner 1995; Kingsley and Bozeman 1997; Kingsley and Farmer 1997; Kingsley, Bozeman, and Coker 1996). A particular advantage of the approach is that it not only provides an indication of the type and amount (though not a single numerical index) of value, but also gives insight into the reasons benefits are achieved. Thus, R&D value mapping (RVM) is useful for policy management strategies seeking to replicate success.

RVM has much in common with earlier case study-based attempts to assess research but is in many respects a significant departure. As in previous case studies of R&D impacts, RVM focuses intensely on particular projects and the events surrounding them. Case studies "tell a story" about the chronology and events contained within the boundaries of the project, and RVM is similar to traditional case studies in that it yields such a narrative. There is also an expectation that the case studies can contain a richness that goes beyond traditional aggregated quantitative studies to provide insights from detail and nuance. RVM,

*The authors gratefully acknowledge support from of the Department of Energy, Basic Energy Sciences and previous contracts from Sandia National Laboratories and the New York State Energy Research & Development Authority. A number of persons have made useful comments on the RVM method. We are particularly grateful to David Roessner, Juan Rogers, Gretchen Jordan and Iran Thomas.

however, seeks to avoid some of the pitfalls of traditional qualitative analysis.

Case studies are faulted as interesting stories, which provide little systematic explanation. The RVM approach, beginning with carefully specified and testable models of causation, as well as a scheme for linking the cases, yields both particularistic and generalizable data. The particularistic information is much like that which is derived from a traditional case. The generalizable data comes from the quantification of elements across cases. Thus, each project "tells a story" and simultaneously gives rise to systematically measured observations.

We do not apply RVM in this paper but instead outline the method and discuss previous and ongoing applications. The objectives of our paper, then, are to articulate the method, compare it to other approaches of R&D impact assessment, and to assess its strengths and weaknesses. We begin by reviewing the use of case studies for understanding the impacts of R&D and technology development.

A Brief History of the Use of Case Study in R&D Impact Analysis

Case study methods, which have been described by some quantitatively oriented social scientists as little better than cultivation of anecdotes (Luukkonen-Gronow 1987), have of late received much attention. Among the many contexts in which case studies have become popular, those in the field of R&D impact analysis have been among the more innovative.

The use of case study methods for evaluations of R&D impacts has been shaped by two research questions: (1) What are the linkages between R&D and economic innovation? (2) Are R&D projects meeting the policy objectives established for the sponsoring organization that mandate linkages between R&D and the economy? Answers to the first question have been the preoccupation of policymakers since World War II when the impact of science on the welfare of the nation became dramatically clear (Ronayne 1984). Answers to the second question have been the preoccupation of industry and government agencies who must demonstrate the economic benefits of specific R&D projects (Roessner 1988).

Initially, case study was employed in the hope of developing concepts and methods that would allow a more precise understanding of terms such as invention, innovation, technology transfer, or basic, applied, and development research (Ronayne 1984). The ultimate thrust of this research was to develop concepts and methods that would allow a more explicit and thoughtful articulation of the causal relationships that link R&D and the economy (Freeman 1977).

There have been four genre of case studies used in the post-World War II era for examining the impacts of R&D. Three are different forms of retrospective analysis: (1) historical descriptions, (2) research event studies, and (3) matched comparisons. The fourth is a combination of retrospective analysis with other methodologies such as aggregate statistics, peer review, bibliometrics, and econometrics (Logsdon & Rubin 1985). Though the development of these case study types are roughly sequential and build upon the failings of earlier studies, the development of new techniques has not resulted in the obsolescence and retirement of the earlier approaches.

The earliest approach was to conduct *historical descriptions* of the development of a specific technology. The work of Jewkes, Sawers, and Stillerman (1969) is an example of this genre of case study, examining the relationship between R&D and innovation by tracing innovations back to fundamental supporting inventions. Similarly, Carter and Williams (1957) examined the stages in the generation and application of scientific knowledge from basic research to the commercial decision of innovation investment. Though historically informative, this approach did not result in a structured analytic framework with well-defined concepts and methods of measurement.

From the 1960s to the 1970s, a series of massive case study projects were sponsored by government agencies in an effort to understand the linkages between R&D and economic growth. Studies such as Project Hindsight (Sherwin & Isenson 1967) sponsored by the Department of Defense, and the Technology in Retrospect and Critical Events in Science project (TRACES) (IIT 1968), sponsored by the National Science Foundation, advanced the analytic techniques used in retrospective analysis by identifying "research events" in the development of specific technologies. Research events are defined as the occurrence of a novel idea and the subsequent period where the idea is explored. Thus, the technique was to take specific research technologies and divide them into the research events that led to the successful development of the technology. Another development in retrospective analysis was to compare innovations that had been determined *a priori* to be of different types. For example, Project SAPPHO (SPRU 1972) conducted pairwise comparisons of innovations that were successes and failures in terms of commercial diffusion.

The empirical results of these studies dramatically conflicted, reflecting the interests of the organizations that had sponsored the studies (Krielkamp 1971; Mowery & Rosenberg 1982); nor did these studies establish a strong conceptual base from which further research could build (Mowery & Rosenberg 1982). Economic and bibliometric techniques began to replace retrospective case studies as the preferred methods to examine the link between R&D and economic innovation (Layton 1977, Luukkonen-Gronow 1987).

Throughout this period, case study had also been used to evaluate the performance of specific R&D projects within the context of a policy objective. These objectives normally have an implicit, or explicit, assumption that R&D directly affects the economy (Roessner 1988), but the goal of the case study emphasizes evaluation of project

performance in preference to developing contributions to theory. Though case studies seeking to establish linkages between R&D and the economy failed in establishing a strong theoretical base, they nonetheless had a significant methodological influence upon case studies emphasizing project evaluations. Evaluation studies have mimicked the retrospective case study designs used to develop theory. For example, a recent case study conducted by Oak Ridge National Laboratory uses a form of retrospective analysis charting the Department of Energy's contribution to the development of several building innovations (Brown, Berry, & Goel 1991).

But the frustration with the findings from case study design also led to a variation in case study research that combines several methodological techniques. As noted above, these multi-method approaches bring together case study with peer review, bibliometric techniques, and econometric analysis under the heading of impact analysis (Logsdon & Rubin 1985). The goal of impact analysis is to look for levels of agreement between the different techniques employed (Nelson 1982; Logsdon & Rubin 1985; Meyer-Krahmer 1988).

Strengths and Weaknesses of Case Study

Yin (1994) has summarized the major strengths and limitations inherent in all case study designs. There are three strengths to case studies. First, this method is very useful for addressing questions regarding how and why a phenomenon behaves. In other words, the findings of case study research reveal a rich detail of information that highlights the critical contingencies that exist among the variables. Second, this method is very useful for exploration of topics when there is not a strong theory to which one can appeal. It is particularly useful for addressing contemporary subjects where there is not a knowledge base to draw upon. Similarly, unlike some quantitative methods, case study is very forgiving to the researchers own learning process of the social phenomenon that is being observed.

Yin (1994) suggests there are three fundamental problems. First is the concern over the lack of rigor of case study research. The thrust of this concern is that the format of case study allows equivocal evidence, or biased views, to influence the directions of the findings and conclusions. This problem grows out of the nature of the data collected, which is often in narrative form and in large volumes of information.

A second problem is that, though case study is useful for ordering information, there is little inherent in the methodology for assessing causality or making scientific generalization. Yin (1994) argues that concerns regarding the lack of rigor of case studies are exaggerated and outlines ways to remedy this criticism. Case study research designs can and do utilize multiple case comparisons (as has been seen in R&D impact evaluations in each of the four quadrants of Table 1). This can strengthen the external validity of studies without sacrificing internal validity. Similarly, researchers develop a framework of analysis for making comparisons across cases with greater rigor than that normally associated with case study.

A third concern regarding case study is that it takes a great deal of time to collect and analyze the data when attempts are made to use case study in a scientific manner addressing the problems of validity and reliability. Relatedly, it is also an expensive method to conduct. The combination of the two reduces the practicality of this method in many research questions.

A variety of innovations have been developed during R&D evaluations to mitigate the weaknesses of case study methods while capitalizing upon the strengths. Kingsley (1993) has analyzed these using a two-dimension typology. The first dimension addresses whether the research question pursues (1) the development of theory relating R&D to social and economic innovation or (2) the evaluation of outcomes in relation to organizational goals. The second dimension distinguishes between the use of case study by the public or private sector. A diagram of this typology, dividing the use of case studies into four quadrants, is provided in Table One. Each quadrant is evaluated according to the following criteria: first, the research question framing the study; second, the case selection method; and third, the analytical methods employed.

Case studies in Quadrant 1 are designed to develop theory regarding the role of the government in supporting R&D. There are relatively few case studies associated with this quadrant. TRACES (ITT 1968) and Project Hindsight (Sherwin & Isenson 1967) are the best known of the genre. This is partially due to the enormity of the task. Those that have been conducted were massive undertakings, requiring years of effort by large teams of researchers (Ronayne 1984). These case studies seek to identify the contributions of R&D to technological innovation with the intention of determining the proper role of government in supporting research. Selection of cases has been determined by the types of technologies with which the sponsoring agency has been working. This type of case study divides each path of technological development into significant research events. The nature of the government's role is described for each of these events.

Case studies in Quadrant II attempt to do a similar type of theory development but apply their efforts to private sector developments of technology. The studies attempt to relate the events leading to a specific technological innovation with the associated industrial structures, organizational structures and managerial practices (Jewkes et al. 1969; Langrish et al. 1972). Histories are developed of technological innovations and matched comparisons are made between those that were successfully commercialized and those that were not. Selection of the cases for study generally are made based upon convenience and access.

Case studies in Quadrant III are conducted as part of a government agency's efforts to evaluate projects that it

Table 1. Typology of the uses for case study for R&D evaluation

	Public Sector R&D	Private Sector R&D
	Quadrant I:	Quadrant II:
Theory-Driven Research	Large-Scale, Research Event Histories	Technology Histories, Matched Comparisons
	Quadrant III:	Quadrant IV:
Evaluation-Driven Research	Social and Economic Impact Analysis	Firm or Industry Impact Analysis

has sponsored (Kerpelman & Fitzsimmons 1985; Logsdon & Rubin 1988). Thus, the research question is narrower, limited to assessing whether the project or program is meeting the policy objective. Generally, studies of this sort have focused upon cases that agency managers have identified as "successful." Three methods are combined with a project development history to conduct the impact analysis: (1) aggregate statistics; (2) production functions; (3) peer review.

Quadrant IV contains private sector evaluations of R&D projects (Levinson 1983; Bard et al. 1988; Utterback et al. 1988). The focus of the evaluation is not directly upon the R&D project but instead upon the commercial performance of the industry or firm. Case selection is generally both opportunistic and focused upon successful projects. The case study usually is comprised of a description of the industry's or the firm's structure, a history of the development of a key technology, and a history of the market for the technology.

The primary strength of these different genres of case studies is that they provide a context for understanding the many contingencies that affect how and why R&D has impacts. However, because the use of case study is constrained in R&D evaluations to a form of retrospective analysis, there has been little progress towards predictive models.

The R&D Value Mapping Approach

In capsule, RVM begins with one or more analytical model that tracks flows of knowledge and specifies possible outcomes of R&D projects. The outcomes are modeled in terms of sequences of events, depicted as a branching model. Each step in the model might be either the final outcome for the project or a preliminary stage to the next step. Thus, the sequences might include the following steps:

(step 1) project completed (yes, no), [if yes...]
(step 2) results disseminated outside the laboratory (yes, no), [if yes...]
(step 3) results used by an individual or organization not affiliated with the lab (yes, no), [if yes...]
(step 4) product developed from results (yes, no), [if yes...]
(step 5) product marketed (yes, no) [if yes...]
(step 6) outcomes [for example, sales from product, or other measures of benefits, costs, and disbenefits].

Measuring the variety of benefits and disbenefits that may result from a project specifies a second dimension of R&D outcomes. As an illustration, Table 2 provides a potential list of the commercial benefits from R&D projects. The potential benefits will, of course, vary according to the objectives of projects. However, there is nothing inherent to the RVM approach that ensures that outcome measures are limited to benefits. For example, Kingsley, Bozeman, and Coker (1996) examined the impacts of failures to transfer technology from R&D projects.

RVM involves measuring a variety of hypothesized project attributes (e.g. resources devoted to a project; the number of industrial participants; disposition of IPR) against the branched patterns of outcomes. By conceiving projects in terms of the progress of their results along certain branched alternatives (the steps given above), it is possible to develop predictive models of the factors related to project outcomes vis-a-vis those possible branched alternatives. Essentially, what factors relate to the ultimate path position, the final step, of the project?

After the analytical models and associated hypotheses have been developed, data gathering in RVM is much the same as for a traditional case study. Case selection is also driven by the criteria relevant to the model. The selection process can lead RVM to an embedded case study research design because there are multiple units of analysis, i.e., R&D projects, stemming from a single institutional setting. However, there is no requirement in the RVM method that cases share a common institutional frame. Data sources include personal interviews, documentary evidence, records and files. The resulting data can, indeed, be fashioned into traditional "thick description" cases.

In addition to the results of traditional case studies, RVM provides an analytical device resembling an empirical explanation in quantitative social science. RVM provides quantitative data from cross-case analysis. In some instances the measurement approach is similar to most quantitative studies. Thus, for each case, indicators are developed for such variables as amount of funding for the

Table 2. Illustrative impact assessment table

Impact Type	High Impact	Some Impact	No Impact
Established new company or joint venture			X
Enhanced company's technical capabilities or know-how	X		
Developed new commercial product			X
Developed new commercial process		X	
Improved existing product			X
Improved existing process		X	
Licensed or patented technology or software		X	
Created new jobs	X		
Set industry standard and standard enabling R&D			X
Influenced company's R&D agenda		X	
Company terminated planned process or product (advantageously)		X	
Provided technical knowledge used by company's suppliers or customers			X
Enhanced human capital and skills at company	X		

project, numbers of personnel devoted to the project and, on the benefit side, such variables as estimated monetary benefits and numbers of personnel receiving advanced training. Somewhat of a departure, however, is the attempt to use dummy variables (i.e., 0,1) to measure qualitative aspects of the cases. Thus, it is possible to quantify such variables as whether the lab's technology transfer office was involved in the project (0=not involved, 1=involved), whether a diffusion plan was developed at the outset of a project (0=developed later or not at all, 1=developed at outset), or whether the results of the project required the user to develop new manufacturing processes (0=not required, 1=required). By combining these variables, both the traditional interval-level variables and the dummy variables for the presence/absence of a project attribute, a series of causally relevant independent variables are developed.

These independent variables are then analyzed in terms of the sequential models developed at the outset. This assessment is made both in terms of the step reached in the branching model and the benefits (or disbenefits) that occur. RVM is similar to other case survey techniques whereby multiple coders score individual cases and resulting scores are subjected to an inter-coder reliability analysis (Bullock & Tubbs 1987; Larsson 1993; Wolf 1993). Case scores are then categorized for pattern-matching both within groups of cases and between case groupings.

The research procedures of RVM can be summarized as follows:

1. Develop sequential, but nonlinear, branching model(s) of the flow of knowledge from research to exhaustive set of outcomes.
2. Develop propositions about causal factors related to those outcomes.
3. Develop indicators of costs and benefits from projects and project-related outcomes.
4. Select cases on the basis of factors specified in model and hypotheses.
5. Gather data on cases.
6. Organize data by writing traditional case studies
7. Develop quantitative database by coding the case studies according to the model variables.
8. Validate data coding conventions (e.g. inter-rater reliability indices).
9. Use resulting quantitative data in connection with models, determining (through contingency analysis or dynamic programming) the relation of independent variables to knowledge flows, project outcomes, and benefits and costs.

An early stage in the application of RVM requires assessing the outcomes of projects. This initial assessment can be provided by program managers, principal investigators or others involved with the project. Table 2 also provides an illustrative assessment table, but as knowledge of the projects develop, more sophisticated assessment techniques become appropriate.

A key to the successful application of RVM is to begin with theory-based models depicting the flow of impacts

Table 3. R&D Impact Assessment Techniques; their Applications and Strengths

Method	Technical Needs	Validity/ Reliability	Summative/ Formative	Resources	Time Needed
Qualitative					
Case study	H	H/L	S/F	H	H
Focus group	M	M/L	S/F	L	L
Peer review	L/M	M/M	S/F	M	L
Content analysis	H	H/H	S	M	M
Mixed					
R&D Value Mapping	H	H/M	S	H	H
Delphi Panels	M	M/M	S/F	M	M
Quantitative					
Bibliometric	H	M/H	S	M	M
CBA/ROI	H	M/M	S	M	M
User Survey and Quest.	H	H/H	S/F	M	M
Benchmark	M	M/M	S/F	M	M
Quasi-Experiment	H	H/H	S	H	H
Forecasting	H	L/M	S/F	L/M	M
Portfolio Analysis	M	H/M	S/F	L	L
Network Analysis	H	H/H	S/F	M/H	H
Input-Output	H	L/H	S/F	M/H	M/L
Operational Audit	L	M/M	F/S	L	M
Systems/Flow Analysis	L	M/M	F	L	L
Indicator Systems	L	H/H	F/S	L	L
Industry Analysis	M	H/M	S/F	L	L
GIS/Diffusion	H/M	H/M	F/S	M/H	H/M

H – High M – Medium L – Low S – Summative F – Formative

from projects. Since RVM is iterative, it is assumed that these models will be revised continuously during the project in order to inculcate learning. For example, in a study of R&D projects sponsored by the New York State Energy Research and Development Authority, two models were employed in describing the sequence of possible outcomes. A technology absorption model identified the stages in the adoption of a technology by the organizations that participated in the R&D contract. The transfer model identified stages in the movement of technology from a R&D project to adoption by external organizations that did not participate.

Prototype Applications of R&D Value Mapping

Bozeman and colleagues developed the fundamental components of RVM in a study of 31 applied research and development projects (Bozeman et al. 1992; Bozeman & Kingsley 1997; Kingsley, Bozeman, & Coker 1996; Kingsley & Farmer 1997; Kingsley 1995) and then further refined the approach by applying it to a set of cases focusing on projects at DOE laboratories (Roessner, Bozeman, Donez, & Schofield 1996; Bozeman & Donez 1996; Roessner 1996). The more recent prototype applications

focused on three cases—Brookhaven National Laboratory Superconducting Wire (Bozeman & Donez 1996); Stanford Linear Accelerator Thin Film Deposition (Roessner, Bozeman, Donez, & Schofield 1996); and Oak Ridge National Laboratory Ceramic Whiskers (Roessner 1996). A more ambitious and large-scale project, focusing on as many as 50 case studies of projects funded by the Department of Energy's Basic Energy Sciences Division, is now underway.

The Methodological Locus of RVM

As with other evaluation approaches, it is important to understand the position of RVM in connection with other available approaches to assessing R&D and technology development impacts. Table 3, adapted from Bozeman, Shapira, and Youtie (1996), provides an approach to "locating" and assessing RVM in connection with other available R&D impact evaluation approaches.

While it is not our purpose in this paper to provide a systematic assessment of the methods used for R&D impacts analysis (for more detailed treatment see Bozeman & Melkers 1994 and Bozeman, Shapira, & Youtie 1996), it is useful to succinctly describe each of the criteria presented in Table Three and to provide our assessment of RVM in connection with those criteria.

In the first place, RVM is one of the few available techniques that is, at the same time, both qualitative and quantitative. Since it requires as input detailed cases, the cases themselves provide a strong qualitative element. However, since the cases are used to develop indicators and to test explicit models using a sequential path analysis, there is inexorably a quantitative element to RVM.

The criterion "technical needs" refers to the degree of technical training and expertise required for performing the method. A disadvantage of RVM is that the technical needs are extremely high, requiring not only case analysis skills but skills in modeling and, importantly, skills in methodological development. Since there is as yet no template for RVM, its application is not in the least mechanical.

"Validity" refers to the power of the method to ascertain the causal relations in hypotheses about program effects; "reliability" refers to test-retest correspondence. Compared to other available approaches, RVM holds great promise with respect to validity. The combination of in-depth analysis and application of systematic (if not invariant) method means that the inferences from RVM analysis are much better grounded than for most techniques. RVM can also be strong from a reliability perspective. By having several coders review the same large body of case studies inter-coder reliability can be statistically assessed.

"Summative" means the evaluation is chiefly for final program effects; "formative" means findings are useful for program improvement in an ongoing program. RVM is useful for both summative and formative evaluation but is best suited to summative evaluation simply because there must be time for project impacts to occur. This is not to say that it is irrelevant to formative analysis—some project impacts can be observed early on. Moreover, it is by its very nature a "learning technique," requiring adjustment and refinement of models and method as more and more is learned about project outputs and impacts.

The major disadvantage of RVM is that it is inherently resource-intensive. The notion of quantitative treatment of cases depends fundamentally upon having a sufficient number of cases (usually at least 30) to permit quantitative analysis and application of inferential statistics. The requirements are not quite so formidable as they might seem given the possibility of mixing in-depth "base cases" with "mini-case" studies that focus only on the particular variables examined in the RVM models. Even under the best of circumstances this is an approach that requires considerable resources. Similarly, the amount of time required for RVM is considerably greater than for most other approaches.

In sum, the RVM approach is best viewed as "high investment-(potentially) high return." It requires considerable resources, considerable technical and methodological expertise (including some receptivity to methodological innovation), but its expense and effort is redeemed by detailed knowledge of cases (as with most case study approaches) as well as a systematic set of explanations of impacts.

References

Averch, H. A. "Economic Approaches to the Evaluation of Research." *Evaluation Review* 18(1), 1994, pp. 77-88.

Bard, J. F., R. Balachandra, and P. E. Kaufmann. "An Interactive Approach to R&D Project Selection and Termination." *IEEE Transactions on Engineering Management* 35, 1988, pp. 139-146.

Bozeman, B., et al. *R&D Impacts at the New York State Energy Research and Development Authority: Final Report.* Syracuse, NY: Center for Technology and Information Policy, Syracuse University, 1992.

Bozeman, B., and F. Donez. "Brookhaven National Laboratory, Superconducting Materials and Magnet Technology." Case study prepared for Sandia National Laboratories, 1996.

Bozeman, B., and J. Melkers (eds.). *Evaluating R&D Impacts: Methods and Practice.* Norwell, MA: Kluwer Publishing, 1993.

Bozeman, B., and J. D. Roessner. *Prototype Case Studies for R&D Value Mapping.* Atlanta: Technology Policy and Assessment Program, Georgia Institute of Technology.

Monograph prepared for Sandia National Laboratories, 1995.

Bozeman, B., P. Shapira, and J. Youtie. "Designs for Evaluating Impacts of the Georgia Research Alliance." Study prepared for the Office of Planning and Budget, State of Georgia, 1996.

Brown, M. A., L. G. Berry, and R. K. Goel. "Guidelines for Successfully Transferring Government-Sponsored Innovations." *Research Policy* 20, 1991, pp. 121-143.

Bullock, R. J., and M. E. Tubbs. The Case Meta-Analysis Method for OD. *Research in Organizational Change and Development* 1, 1987, pp. 171-228.

Carter, C. F., and B. R. Williams. *Industry and Technical Progress: Factors Governing the Speed of Application of Science to Industry.* London: Oxford University Press, 1957.

Freeman, C. "Economics of Research and Development." In *Science, Technology and Society: A Cross-Disciplinary Perspective,* ed. I. Spiegel-Rosing and D. De Solla Price. London: Sage Publications, 1977.

Illinois Institute of Technology Research Institute. *Technology in Retrospect and Critical Events in Science.* Washington, DC: National Science Foundation, 1968.

Jewkes, J., D. Sawers, and R. Stillerman. *The Sources of Invention, 2nd edition.* New York: W.W. Norton & Co., 1969.

Kerpelman, L. C., and S. J. Fitzsimmons. *Methods for the Strategic Evaluations of Research Programs: The State of the Art.* Washington, DC: National Science Foundation, 1985.

Kingsley, G. "The Use of Case Studies in R&D Impact Evaluations." In *Evaluating R&D Impacts: Methods and Practice*, ed. B. Bozeman and J. Melkers. Norwell, MA: Kluwer Publishing, 1993, pp. 17-42.

Kingsley, G., and B. Bozeman. "Charting the Routes to Commercialization: The Absorption and Transfer of Energy Conservation Technologies." *International Journal of Global Energy Issues* 9, 1997, pp. 8-15.

Kingsley, G., and M. C. Farmer. "Using Technology Absorption as an Evaluation Criterion: The Case of a State R&D Program." School of Public Policy Working Paper 97-02. Georgia Institute of Technology, Atlanta, GA, 1997.

Kingsley, G., B. Bozeman, and K. Coker. "Technology Transfer and Absorption: An 'R&D Value-Mapping' Approach to Evaluation." *Research Policy* 25, 1996, pp. 967-995.

Kreilkamp, K. "Hindsight and the Real World of Science Policy." *Science Studies* 1, 1971, pp. 43-66.

Langrish, J., M. Gibbons, W. G. Evans, and F. R. Jevons. *Wealth from Knowledge*. London: MacMillan, 1972.

Larsson, R. "Case Survey Methodology: Quantitative Analysis of Patterns across Case Studies." *Academy of Management Journal* 36(6), 1993, pp. 1515-1546.

Layton, E. "Conditions of Technological Development." In *Science, Technology and Society: A Cross-Disciplinary Perspective,* ed. I. Spiegel-Rosing and D. De Solla Price. London: Sage Publications, 1977.

Logsdon, J. M., and C. B. Rubin. *An Overview Of Federal Research Evaluation Activities*. Washington, DC: National Science Foundation, 1985.

Luukkonen-Gronow, T. "Scientific Research Evaluation: A Review of Methods and Various Contexts for Their Application." *R&D Management* 17, 1987, pp. 207-221.

Meyer-Krahmer, F. "Evaluation of Industrial Innovation Policy: Concepts, Methods and Lessons." In *Government Innovation Policy: Design, Implementation, Evaluation*, ed. J. D. Roessner. New York: St. Martin's Press, 1988.

Mowery, D. C., and N. Rosenberg. "The Influence of Market Demand upon Innovation: A Critical Review of Some Recent Empirical Studies." In *Inside the Black Box: Technology and Economics*, ed. N. Rosenberg. New York: Cambridge University Press, 1982.

Nelson, R. R. "Government Stimulus of Technological Progress: Lessons from American History." In *Government and Technical Progress: A Cross-Industry Analysis*, ed. R. R. Nelson. New York: Pergamon Press, 1982.

Rao, I. K. "Strategies for Best Practice in Research and Technology Institutes—An Overview of a Benchmarking Exercise." *Scientometrics* 35(2), February 1996, pp. 265-270.

Roessner, J. D. "Whisker-Reinforced Ceramics at Oak Ridge National Laboratory," Case study prepared for Sandia National Laboratories, 1996.

Roessner, J. D. *Government Innovation Policy: Design, Implementation, Evaluation.* New York: St. Martin's Press, 1988.

Roessner, J. D., B. Bozeman, F. Donez, and T. Schofield. "Understanding Thin Film Deposition at the Stanford

Syncrotron Radiation Laboratory." Case study prepared for Sandia National Laboratories, 1996.

Roessner, J. D., and T. Schofield. "Understanding Thin Film Deposition at the Stanford Syncrotron Radiation Laboratory." Case study prepared for Sandia National Laboratories, 1996.

Ronayne, J. *Science in Government*. London: Edward Arnold (Publishers) Ltd, 1984.

Rush, H., M. Hobday, J. Bessant, and E. Arnold. "Strategies for Best Practice in Research and Technology Institutes—An Overview of a Benchmarking Exercise." *R&D Management* 25(1), 1995, pp. 17-31.

Science Policy Research Unit (SPRU). *Success and Failure in Industrial Innovation*. London: Center for the Study of Industrial Innovation, 1972.

Sherwin, C. W., and R. S. Isenson. "Project Hindsight: Defense Department Study of the Utility of Research." *Science* 156, 1967, pp. 1571-1577.

Utterback, J. M., M. Meyer, E. Roberts, and G. Reitberger. "Technology and Industrial Innovation in Sweden: A Study of Technology-Based Firms Formed Between 1965 and 1980." *Research Policy* 17, 1988, 15-26.

Wolf, P. J. "A Case Survey of Bureaucratic Effectiveness in U.S. Cabinet Agencies: Preliminary Results." *Journal of Public Administration Research and Theory* 3(2), 1993, pp. 161-181.

Yin, R. K. *Case Study Research: Design and Methods, 2nd Edition*. Newbury Park, CA: Sage Publications, 1994.

Author Biographies

Barry Bozeman is Professor of Public Policy and Director, School of Public Policy, Georgia Tech. His most recent book, written with Michael Crow, is *Limited by Design: Chaos in the U.S. R&D Laboratory System* (Columbia University Press, to appear 1998).

Gordon Kingsley is Assistant Professor of Public Policy, Georgia Tech. He teaches courses in organization theory and technology policy. His research focuses on technology transfer and evaluating R&D impacts.

Tracking Customer Progress: A Follow-up Study of Customers of the Georgia Manufacturing Extension Alliance*

Jan Youtie

Economic Development Institute, Georgia Institute of Technology
Atlanta, GA 30332

Philip Shapira

School of Public Policy, Georgia Institute of Technology
Atlanta, GA 30332

Abstract

Time lags often exist before the economic impacts of technology promotion programs fully materialize. For one manufacturing technology deployment program, the Georgia Manufacturing Extension Alliance, this study gathered expected impact data soon after the point of service. Customers were then surveyed one year later and asked about impacts actually realized. A comparison showed that for the average project, actual benefits reported at the one-year survey mark were generally lower than benefits expected immediately after project completion, while actual costs were generally higher than expected costs. For high performing projects, however, the study found that actual benefits after one year were substantially higher than the benefits initially expected soon after assistance was completed. This study explores the implications of these findings for technology program evaluation and methods of performance measurement.

Introduction

Federal and state governments have made extensive investments in policies to promote technology development and deployment by the business sector. Programs have been established in a variety of technology promotion areas, including support for start-up technology ventures, collaborative research and development between firms and public technology institutions, the transfer and commercialization of technologies developed by universities and federal laboratories, and the deployment of new manufacturing technologies in industry. One recent study estimates that combined federal and state support for such cooperative technology programs now exceeds $3 billion annually (Berglund and Coburn 1995). At the same time, demands on these programs to demonstrate economic and business results have also grown, in parallel not only with increased budgets but also with the renewed interest throughout the public sector in the last few years in performance measurement and the more effective delivery of public services (Carlisle 1997; Gore 1993).

However, technology promotion programs have specific characteristics that often make it difficult to present hard evidence which can attribute economic and business outcomes to publicly supported inputs. For example, technology promotion programs frequently provide assistance or resources that require additional downstream private actions and investments for results to materialize. Public policy can usually only encourage these further private steps, and cannot control them. Moreover, it generally takes a length of time before program assistance is translated into action by assisted businesses and, in turn, translated into realized effects on production, sales, or jobs (Bozeman et al. 1997). Policy-makers recognize this when they affirm that technology promotion programs are long-term initiatives that cannot be expected to show significant results over short time frames. Yet these same policy-makers also desire evidence of rapid impacts as they make annual budget decisions (Shapira et al. 1996). This is a wish that almost all program managers aim to fulfill as they try to justify the economic value they believe their program has generated.

One strategy that technology programs use to present at least some immediate information about long-run effects is to estimate or project program benefits and outcomes very soon after assistance has been completed. Companies receiving assistance from technology promotion programs may be asked to report anticipated figures for sales increases or new jobs created as a consequence of program assistance. Program staff or outside evaluators may then report the anticipated results directly to sponsors or may construct some type of model which corrects for or extrapolates based on the fact that the results are "anticipated" and not "actual" results (Pressman 1996).

*The authors would like to acknowledge the assistance of Alan Olisko and Mike Lane for their assistance in conducting the surveys. The comments of Gretchen Jordan and John Mortensen were also most helpful in improving the paper.

The problem is that anticipated results may not be valid indicators of actual program impacts. Even with hindsight, it is difficult for companies to put precise dollar values on the impacts of program assistance. Detailed records are rarely kept, while in many cases technology programs influence "soft" factors such as know-how, skill, or knowledge flows which, if not entirely intangible, are complex to monetarize. The problems of estimating the economic value of assistance are even greater when companies are asked to project forward. Survey respondents have special difficulties in answering hypothetical questions, particularly about future effects, and tend to provide speculative answers (Converse and Presser 1990; Smith 1981). If too much time elapses between program participation and surveying, response rates and the ability of customers to provide data about the impacts of program participation may decline as personnel change or other business events occur. As with most evaluation designs, given limited resources (and the limited patience of customers in responding to repeated requests for information), trade-offs are involved in selecting not only what questions should be asked, but at what point in time those questions should be administered. Other evaluation strategies (such as using control groups of non-customers) can provide additional reference points to interpret data from customers, but even with more intricate evaluation designs, the issue of survey timing remains a critical element.

This article analyzes customer reports of impact from a particular technology program—the Georgia Manufacturing Extension Alliance. Using survey data for the same firms collected at two points in time—immediately after program participation and one year later—we are able to examine the reported economic effects of the program and explore the relationships between customer reports of impact and the timing of data collection. We discern that program participation has significant economic effects but we also find that that close to the point of service delivery, customers receiving assistance tend to overestimate the benefits of program participation and underestimate the commitment and resources necessary to achieve the benefits. Subsequent measurement, at about a year after program participation, indicates that customers can provide a more realistic assessment of benefits and costs, although with some drop-off in response rates. The one-year survey shows that program participants still receive net benefits, although at a lower level than first anticipated immediately after the close of the project. Importantly, for a relatively small number of cases where program participation results in very large positive impacts, we find some evidence that immediate post-project measurements underestimate the scale of the ensuing outcomes. In following sections, these findings and their background are discussed in more detail.

Program Context

The Georgia Manufacturing Extension Alliance (GMEA) provides industrial extension and technology deployment services to the state's manufacturers. GMEA's services are focused particularly towards the small and medium-sized companies with fewer than 500 employees that comprise the vast majority of Georgia's 10,000 manufacturers. The lead organization in GMEA is Georgia Institute of Technology's Economic Development Institute, which builds on a 30-year history of industrial extension service provision. The Economic Development Institute operates a network of 18 regional field offices, staffed with industrially experienced engineers and business professionals. Field office services are supported by staff in several program skill centers in such areas as quality, manufacturing information technology, human resource development, strategic management assistance, and energy and environmental services. As necessary, GMEA links companies with specialized expertise at Georgia Tech, federal laboratories, industry technology centers, and private consultants. GMEA also works in partnership with organizations including small business development centers, technical colleges, and utilities to offer a comprehensive array of technology and business support services to firms.

In 1994, GMEA was formed from what was then the state-sponsored Georgia Tech Industrial Extension Service, becoming part of the national Manufacturing Extension Partnership (MEP). Coordinated by the U.S. Department of Commerce's National Institute of Standards and Technology, the MEP is comprised of more than 70 industrial extension and manufacturing technology deployment programs operating in all 50 states. The MEP itself is a partnership involving federal, state, and industry resources. Additional federal resources provided through the MEP have allowed GMEA to increase the scale of its operations and forge new linkages with state, federal, and industry groups. The MEP has encouraged the development of systematic evaluation procedures for its manufacturing extension affiliates, including the implementation of standardized performance measures, periodic reporting, and customer surveying (National Institute of Standards and Technology 1994; National Institute of Standards and Technology 1996).

Through company assessments, formal projects, informal assistance engagements, training workshops, technology demonstrations, and other activities, GMEA now serves about 1,000 Georgia manufacturers annually. To understand and assess the impacts and outcomes of these services to manufacturers, GMEA has established an explicit evaluation protocol (Shapira and Youtie 1994). This evaluation protocol assesses program and customer impacts through several complementary methods. The first is a post-project survey of customers 30 to 45 days after a project has been closed. This survey asks for customer satisfaction information and reports of both received and anticipated quantitative and qualitative outcomes. The second is a one-year follow-up telephone survey after the first year to further estimate actual (not anticipated) out-

comes. In addition to the two customer surveys, the GMEA evaluation strategy also includes a controlled survey sent every two years to all Georgia manufacturers with 10 or more employees designed to assess longer-term impacts of the program, as well as cost-benefit analyses of the program and case studies of successful projects.

Study Methodology

In this article, we focus on the results of the first two methods: the post-project survey of customers and the one-year follow-up survey. The post-project survey procedure was instituted by GMEA in 1994. Each month, information about closed projects with manufacturers was drawn from the program's management information system. For projects with significant program intervention (defined as eight hours or more of program staff assistance), a standardized satisfaction and impact questionnaire was sent centrally by mail to the principal company contact for the project. Shorter program interactions with companies, such as initial visits or informal consultations, were not formally evaluated through this procedure. In 1994, about 55 percent of the program's interactions with customers were for 8 hours or more (by 1997, these more lengthy interactions had grown to represent two-thirds of program interventions). The time required for information reporting and mailing meant that customers usually receive the post-project questionnaire about 30-45 days from the completion of the project. As necessary, the first questionnaire is followed by a second mailing and telephone contact. The response rate to the post-project survey procedure was relatively high—about 70 percent.

In July and August of 1995, we then conducted a telephone follow-up survey of the first wave of completed 1994 GMEA projects. Initially, there were 129 projects for which post-service mail questionnaires were available. Sixteen of these projects were excluded from further analysis because of various reasons (for example, the projects were duplicates or it turned out that the projects were still ongoing). This left a database of 113 projects. Eighty percent of these projects were at least a year old. (The remaining projects were generally at least nine months old, although one was less than nine months old.) Since most of the projects surveyed were at least a year old, we refer to this follow-up survey as the one-year follow-up survey. Customer contacts for 75 of these 113 projects were reached during the one-year follow-up survey administration period.

The overall response rate to the follow-up survey was 66 percent. The primary reason for non-response (for 28 of 38 non-respondents) was that the company did not return telephone calls or faxes, with no further information available. In other cases, the company was reached but the principal project contact had left, the company declined to respond to the questionnaire, or discrepancies were discovered in the initial database. We further explored the characteristics associated with non-response by examining differences between respondent and non-respondent answers in the post-project mail questionnaire (Table 1). No clear direction of bias emerged. Respondents were somewhat more likely to anticipate taking action as a result of the assistance and services they received than were non-respondents. At the post-project stage, 85 percent of subsequent follow-up survey respondents anticipated taking action, while 75 percent of the follow-up survey non-respondents anticipated taking action ($p=.151$). On the other hand, at the post-project stage, non-respondents were somewhat more satisfied with the assistance and services they received than were respondents. The mean overall service rating on the post-service questionnaire was 4.2 for one year follow-up survey respondents compared to 4.4 for non-respondents ($p=.131$). These ratings were based on a five-point scale in which 1=poor, 3=adequate, and 5=excellent.

Additionally, the number of workers employed in respondent facilities was compared to that for non-respondents, as well as to the total GMEA customer base. GMEA customers that received surveys tended to be larger than the general GMEA customer pool (perhaps because very small firms are served through means other than formal projects, including informal assistance, workshops and seminars). One year follow-up survey respondents tended to employ fewer employees than non-respondents, although the differences were not significant.

Program Results Reported in the Customer Surveys

The business and economic impacts of a program like GMEA are determined by a sequence of events. First, does the customer take any action as a result of the assistance and services provided? Second, what are those actions and what impacts do they have in terms of sales, cost savings, capital investment, and jobs? We now turn to probe these questions, drawing on the information and results reported by GMEA customers in the two surveys. Our aim is to determine the likelihood and scale of reported actions and impacts and to make comparisons between customer responses to the one-year follow-up survey and those provided in the post-project mail questionnaire. (We leave for analysis in subsequent articles such questions as how the business and economic performance of assisted firms compares, over the long-run, with non-assisted controls.)

Taking Action

In the post-project mail questionnaire, 64 customers (85 percent) took or anticipated taking action as a result of the assistance and services they received. One year later, 51 of these firms (or 68 percent) actually took action (Table 2). Additionally, one year later, nine projects were reported to be on hold (i.e., the firm was still considering whether to

Table 1. Comparison of respondents and non-respondents to GMEA one-year follow-up survey

	One-year follow-up survey	
	Respondents	Non-respondents
Response to one-year follow-up survey		
Number	75	28
Percent of total (n=113)	66.40%	33.60%
Customer employment size (a)		
Mean	226	331
Median	84	140
Response to post-project survey		
Overall project rating (b)	4.20	4.45
Taking action anticipated, percent	85.30	75.00

Source: Analysis of post-project surveys of 113 GMEA projects closed in 1994 and subsequent responses to 1995 one-year follow-up survey.

Notes:

a. In September 1995, the mean number of employees in the total GMEA customer base was 189; the median was 57.

b. Rating based on five-point scale in which 1=poor; 3=adequate; and 5=excellent.

implement the project recommendations). It appears that one year after, customers did not take action to the extent they thought 30-45 days after project closure. This is due mainly to decisions to put some projects on hold, rather than a decision to definitely not take action on project recommendations. If some of the projects reported to be on hold one year after project close-out are actually implemented, there will be a narrowing of the gap between the follow-up survey rate of action and the post-project survey expectation.

Change in Annualized Sales

In the post-project mail questionnaire, 30 percent of the project contacts anticipated sales increases as a direct result of GMEA assistance and services. One year later, only 17 percent actually experienced sales increases. Actual median sales increases were lower than anticipated median sales increases. For a few outlying customers, however, actual sales increases were substantially higher than anticipated sales increases; the high impact outlying customer reports positively skewed the mean such that actual mean sales increases exceeded anticipated mean sales increases. It is possible that these very large sales increases are due to factors not entirely attributable to program intervention. To err on the side of conservatism in attributing program intervention to sales impacts, we thus also report an adjusted mean, which excludes outliers more than three standard deviations from the mean. The adjusted mean annualized sales increase for those companies reporting this outcome was $207,000 in the one-year follow-up, compared with the initial prediction of $171,000 in the immediate post-project survey. (See Table 2.)

Changes in Annualized Operating Costs

Technology deployment projects can result in changes in operating costs as program staff help firms to better use labor or make savings in factors such as materials or energy. Based on the one-year follow-up survey, Table 3 shows the proportion of GMEA projects which led to changes in labor, material, waste minimization, energy, and other areas. On one-fifth of the projects, companies reported labor operating cost changes resulting from GMEA project participation. Five projects actually produced higher operating costs following the creation of new jobs, resulting in mean added annualized costs of $25,000 across all respondents. In just under one-fifth of the projects, waste minimization savings were reported, with mean annualized savings of $27,000 for reporting companies.

There were two main differences between the post-service survey and the one-year follow-up survey in terms of how questions about operating costs were asked. The post-service survey had one general question about operating costs, but the one-year follow-up survey had several questions asking about each operating cost component individually. Furthermore, the post-service survey only asked about savings, whereas the one-year follow-up survey asked about additional costs as well as additional savings.

Table 2. Comparison of business reported impacts of GMEA project assistance using post-project and one-year follow-up surveys

Impact categories	Post-project survey			One-year follow-up		
	Number	Percent	Value ($ thousands)	Number	Percent	Value ($ thousands)
Customer action						
Taking action	64	85.3	-	51	68	-
On hold	n/a	-	-	9	20	-
Not taking action	10	13.3	-	15	12	-
Sales increase (annualized)	23	30.7	-	13	17.3	-
Mean	-	-	1311.5	-	-	2689.6
Adjusted mean (a)	-	-	170.8	-	-	206.8
Median	-	-	100	-	-	80
Operating costs (annualized)	34	45.3		30	40	-
Mean	-	-	64.4	-	-	124
Adjusted mean (b)	-	-	n/a	-	-	17.5
Median	-	-	50	-	-	20
New capital expenditures	24	32	-	21	28	-
Mean	-	-	272.2	-	-	407.3
Adjusted (c)	-	-	57.1	-	-	244.5
Capped mean (d)	-	-	116	-	-	207.3
Capped adjusted mean (d)	-	-	57.1	-	-	165.6
Median	-	-	25	-	-	87.5
Capital expenditures avoided	13	17.3	-	7	9.3	-
Mean	-	-	83	-	-	74.4
Median	-	-	50	-	-	35

Source: Analysis of post-project surveys of 75 GMEA business customers with projects closed in 1994 who responded to one-year follow-up survey. Post-project survey conducted 30-45 days after project closed. One-year follow-up survey conducted in July 1995.

Notes:
a. Adjusted mean excludes $30 million sales impact reported for one project which was more than three standard deviations from the mean.
b. Adjusted mean excludes $2.1 million operating cost savings reported for one project which was more than three standard deviations from the mean.
c. Adjusted mean excludes $3.5 million capital expenditure reported for one project which was more than three standard deviations from the mean.
d. Capital expenditures capped at $1 million.

Table 3. Changes in operating costs reported by companies as a result of GMEA project assistance, one-year follow-up survey

Operating cost items	Companies reporting impacts		Value of reported impacts $ thousands	
	Number	Percent	Mean	Median
Labor cost impacts	16	21	+25	+25.0
Material cost impacts	8	10.7	-10.8 (a)	-40
Energy cost impacts	8	10.7	-58.3	-60
Waste minimization cost impacts	14	18.7	-26.6	-10
Other cost factors	7	9.3	-27.5	-22.5

Source: Analysis of one-year follow-up survey of 75 GMEA projects, July 1995.
Note:
a. Adjusted mean reported for material cost impacts, excluding $2 million materials savings reported for one project which was more than three standard deviations from the mean. Unadjusted mean was $408.6 thousand.

Despite these differences, some comparison of operating cost changes between the post-project and one-year surveys can be made. Overall, in the one-year survey, 40 percent of the customers said that GMEA projects had led to at least one of the operating cost items. This is slightly less than the 45 percent of customers anticipating operating cost impacts in the immediate post-project survey (See Table 2.) Actual annualized median cost savings ($20,000) were lower than anticipated annualized median cost savings ($50,000). Again, actual mean cost savings exceeded anticipated mean cost savings, because a few high impact projects had an upward influence on the one-year follow-up survey mean.

Capital Expenditures

Capital expenditures include investments in plant, equipment, or other capital items. Project respondents generally viewed capital expenditures as being for equipment, although one respondent referred to construction of a new facility as a capital expenditure. Two aspects of capital expenditures were addressed: capital investments made and capital expenditures avoided.

Substantially more projects resulted in new capital investments than in capital expenditures avoided. Twenty-eight percent of the projects in the one-year follow-up survey effort led to increases in capital expenditures, a rate fairly close to the 32 percent in the post-service mail questionnaire anticipating capital expenditures. Only nine percent of the projects actually helped companies avoid capital expenditures according to the one-year follow-up survey, although 17 percent of the projects were anticipated to help avoid such expenditures in the immediate post-project survey.

Where capital investments were made, they were significant in monetary terms. In the one-year survey, customers reported that their capital investment-related projects resulted in mean expenditures of over $400,000. This mean was bolstered by a few very large capital investments. It may not be reasonable to attribute these unusually large investments entirely to GMEA actions, although it may be fair to attribute a fraction (Shapira and Youtie 1995). The average dropped to $165,600 when one very large outlying project was excluded, and expenditures related to another project capped at $1 million. Still, even this capped adjusted mean represents a significant level of investment.

Capital expenditures avoided had smaller dollar impacts than capital investments made. In the one-year follow-up, companies that reported avoiding capital expenditures due to project assistance reported mean savings of $74,000.

When the results of the post-project and one-year surveys are compared, we find what is by now the consistent pattern. When they are made, the capital investments required to implement project recommendations turn out to be higher at the one-year mark than anticipated 30-45 days after project completion. Equally, when they occur, avoided capital expenditures reported in the one-year follow-up were, on average, lower than originally anticipated in the post-service mail questionnaire.

Employment Impacts

Technology deployment projects may help manufacturers create new jobs or save existing jobs. However, through changes in technology or manufacturing operations, these projects might lead to fewer jobs in some cases. The post-project mail questionnaire asked about new jobs created or current jobs saved, but did not ask about jobs lost. This omission was rectified in the one-year follow-up survey.

In the one-year survey, we found that more companies had added jobs than they anticipated at the point of service

Table 4. Business reported employment impacts resulting from GMEA project assistance

	Companies reporting impacts		Reported employment impacts	
	Number	Percent	Mean (jobs)	Median (jobs)
Jobs Added				
Post-project survey	10	13.3	5	2
One-year follow-up	15	20	11	3
Jobs Lost				
Post-project survey	n/a	n/a	n/a	n/a
One-year follow-up	2	2.7	-5	-5
Jobs Saved				
Post-project survey	10	13.3	7	4
One-year follow-up	14	18.7	9	4

Source: See Table 2.

(Table 4). Fifteen projects in the one-year follow-up survey had added jobs, whereas new jobs were anticipated for only 10 projects in the post-project survey. The mean number of jobs added was also higher in the one year follow-up survey than the post-service mail survey, though the median number of new jobs created was only one more (indicating that most of the extra jobs were clustered in a few cases). In the one-year follow-up, two companies reported that jobs had been lost, with a mean of five jobs lost. Overall, the cases and number of jobs added far outweighed the instances of job loss. In the one year survey, fourteen companies reported that jobs had been saved, representing a somewhat higher number of cases than originally anticipated in the post-project survey.

Company Time Commitment

In addition to capital expenditures, companies incur other costs in participating in projects, including the value of committed staff time. Both surveys asked companies to report the total days of staff time committed to the GMEA project. Forty percent of the companies participating in the one-year follow-up survey provided this information. On average, these companies actually committed more days to a project than anticipated (Table 5). In the one-year follow-up, the mean project required more than 266 days, compared with 132 days anticipated in the post-service mail questionnaire. In the one-year survey, the median number of days was 95, in contrast to the 25 days expected at project completion. In many cases, more than one company employee worked on the GMEA project, which helps to account for these rather large reported company time commitments. But—as the differences between the means and medians suggest—there is also a wide upward tilt in the distribution of survey responses. Indeed, in the one-year survey, 15 companies reported that they committed 100 or more days to projects. Complementary data gathered from on-site visits to several customer facilities suggested that GMEA customers were attributing all the days spent on the project from the company's perspective, including personnel commitments before GMEA was even contacted. To adjust (albeit arbitrarily) for the effect of these few large outliers, we also present a capped mean company staff time commitment, capping company staff time attributable to GMEA participation to 100 days. The capped mean company staff commitment in the one-year follow-up remains significantly larger than the company time commitment initially expected in the post-project survey.

High Performing Cases

In several of the impact areas associated with GMEA project participation (e.g., sales impacts and cost savings), the mean one-year impacts exceeded mean end-of-project impacts, but the reverse was true of the medians. This trend suggests that reports of actual impacts are likely to include outlying, high performing projects. We investigated this trend by conducting on-site case studies of two high performing projects. These case studies further illuminated the extent of differences between anticipated and actual impacts (Table 6). While we generally find that companies overestimate benefits and underestimate costs at project completion compared with the results they subsequently report at the one-year mark, the reverse is true for the small number of high performing cases. We found that in two high performing cases, actual sales and jobs impacts one-year out were higher than initially anticipated. A product development project actually yielded $2 million in extra

Table 5. Business reported staff time allocated to GMEA projects

Company staff days on project	Mean (days)	Capped Mean (days) (a)	Median (days)
Post-project survey	132	64	25
One-year follow-up	266	95	95

Source: See Table 2.
Note:
a. Days are capped at 100.

sales bookings and 10 new jobs, substantially more than the $50,000 and six new jobs anticipated in the post-project mail survey. A plant layout project (in which a computer-generated layout was used as a sales tool) generated an $8 million sales increase and 16 new jobs, rather than the $2 million sales increase and 10 new jobs that the company had first anticipated. In addition, the plant layout project yielded substantial operating cost savings (resulting from reduced overtime pay, energy consumption, and scrap rate) that had neither materialized nor had been anticipated at the time of the post-project customer survey.

Conclusion

The one-year follow-up survey suggests that for the average project, GMEA customers tended to overestimate their benefits and underestimate their costs in their anticipated responses soon after the point of service. One mitigating factor is that the one-year time frame may still not be sufficient to allow all benefits to materialize (nor perhaps to allow all costs to be recognized). As part of our longer-term controlled evaluation design, a statewide manufacturing technology survey conducted towards the end of 1996 may provide sufficient records to allow us to track assisted companies over a two-year time frame.

It could be argued that while all companies find it hard to accurately predict project impacts ahead of time, smaller companies with less developed accounting systems may face particular problems. We explored this idea by examining differences between anticipated and actual impacts among large manufacturers (100 or more employees) and small manufacturers (less than 100 employees). No relationship between company size and nearness of anticipated to actual impacts emerged.

Should technology programs eschew immediate post-project impact data entirely in favor of longer term analyses? To be considered here is the fact that many technology assistance programs now survey customers at the point of project completion as part of their quality control procedures. These programs want to assess company satisfaction with services received and obtain rapid feedback on any problems or further needs so as to respond rapidly. To this useful procedure, questions about received and anticipated impacts are often added with little extra marginal cost (since a post-project customer survey is being undertaken anyway). The question thus becomes should programs add a further survey point, some distance away from the completion of a project, to more accurately capture benefits and costs? An additional survey point adds cost, of course, in an environment where financial resources for systematic evaluation efforts are scarce. Additionally, programs are limited in the number of times they can return to their customer to request estimates of program impacts. At some point, too many data requests become a burden to the customer that may discourage further program participation (Shapira et al. 1996). Not to be forgotten is the fact that while subsequent measurements may be more valid, they also may show a smaller level of net program impact for typical customers. In the topsy-turvy world of budgetary politics, at least some program managers may gamble that better information is not worthwhile.

This said, we would argue that it is critically important to conduct additional follow-up studies of program impacts over time, beyond the immediate close-out of a project. The results from post-project surveys are not radically out-of-line with the results reported one year later (and, in fact, for capital expenditures, the match is quite close). If no other survey can be undertaken, a well-managed post-project survey can provide some useful insights. However, at least in the case of technology deployment, a one-year perspective on project impacts allows a rather more valid analysis of realized results than possible through short-term post-project surveys. Of course, within the wise use of resources and the constraints of company forbearance and record-keeping, we would argue that even longer-term customer tracking is desirable (and for other types of technology promotion programs, particularly those which are more research-intensive, long-run follow-ups over multiple years would seem to be essential). Even limited to a one-year follow-up, an added validity is provided which gives policy-makers data and analysis in which they can have a higher level of confidence. While we find, in the case of GMEA at the one-year mark, that program participation benefits are smaller than initially expected and costs are larger, overall the net economic impacts are significant and positive. One-year employment impacts also appear to be higher than initially reported. Finally, we note that subsequent follow-up is needed to properly identify high performing projects. While often not recognized until a period of time has elapsed, the actual economic impacts from these projects can far exceed what was originally anticipated. Moreover, the benefits from high performing projects are such that

Table 6. High performing GMEA cases: Comparison of impacts reported in post-project survey with on-site case study

Project	Post-project survey: Impacts reported	On-site case study: Impacts reported
Product development case		
Sales increases	$50,000	$2 million
New jobs created	6	10
Plant layout case		
Sales increases	$2 million	$8 million
Inventory savings	$750,000	$750,000
New jobs	10	16
Operating cost savings	-	$100,000

Source: On-site case studies conducted in 1995 and analysis of post-project surveys.

they may justify the program by themselves. Without longer-term customer tracking, these particular impacts might be missed.

References

Berglund, D., and C. Coburn. *Partnerships: A Compendium of State and Federal Cooperative Technology Programs*. Columbus, OH: Batelle Press, 1995.

Bozeman, B., J. Youtie, and P. Shapira. *A Design for the Evaluation of the Impacts of the Georgia Research Alliance*. Atlanta, GA: School of Public Policy and Economic Development Institute, Georgia Institute of Technology, 1997.

Carlisle, R. "The Influence of Evaluation on Technology Policy-making and Program Justification: A State-Level Perspective." In *Manufacturing Modernization: Learning from Evaluation Practices and Results,* edited by P. Shapira and J. Youtie. Atlanta, GA: School of Public Policy and Economic Development Institute, Georgia Institute of Technology, 1997, pp. 119-126.

Converse, J., and S. Presser. *Survey Questions: Handcrafting the Standardized Questionnaire.* Newbury Park, CA: Sage Publications, 1990.

Gore, A. *Creating a Government that Works Better and Costs Less: A Report of the National Performance Review*. New York: Plume, 1993.

National Institute of Standards and Technology. *Setting Priorities and Measuring Results.* Gaithersburg, MD: U.S. Department of Commerce, 1994.

National Institute of Standards and Technology. *NIST MEP Reporting Guidelines, Version 2.1,* Gaithersburg, MD: U.S. Department of Commerce, 1996. Electronic copy available by <<http://www.mep.nist.gov/centers/programs/evaluation/guidelines/index.html>>.

Pressman L. "Economic Impact of Technology Licensing: Measuring Induced Investment." Presentation at Workshop *Technology Transfer: How Do We Know What Works?* Santa Fe, NM, April 28-May 2, 1996.

Shapira, P., and J. Youtie, *Georgia Manufacturing Extension Alliance: Overview of the Evaluation Plan.* GMEA Evaluation Working Paper 9401. Atlanta, GA: Georgia Institute of Technology, School of Public Policy and Economic Development Institute, 1994. Electronic copy available through <<http://www.cherry.gatech.edu/mod>>.

Shapira, P., J. Youtie, and J. D. Roessner. "Current Practices in the Evaluation of Industrial Extension Programs." *Research Policy* 25, 1996, pp. 185-214.

Shapira, P., and J. Youtie. *Assessing GMEA's Economic Impacts: Towards a Benefit-Cost Methodology*. GMEA Evaluation Working Paper. Atlanta, GA: Georgia Institute of Technology, School of Public Policy and Economic Development Institute, 1995.

Smith, T., 1981. "Qualifications to generalized absolutes: 'approval of hitting' questions on the GSS." *Public Opinion Quarterly* 45, 1981, pp. 224-230.

Author Biographies

Jan Youtie is a senior research associate in the Economic Development Institute, Georgia Institute of Tech-

nology. Her research interests include economic development, technology policy, evaluation, and market analysis.

Philip Shapira is an associate professor in the School of Public Policy, Georgia Institute of Technology. He teaches and conducts research in the areas of technology policy, regional economic development, and industrial competitiveness.

Evaluating Prediction of Technology Transfer Success: An Interim Evaluation of the Dutch Sensor Technology Program

Frans C. H. D. van den Beemt
Technology Foundation
Utrecht, The Netherlands 3502 GA

Abstract

Technology Foundation STW, a grant organization in the Netherlands, selects research proposals from universities on the basis of their scientific quality and their utilization potential. The proposals are in the field of applied science. STW also assists the research groups in the four years after the grant by bringing together the researchers and the potential users in half-day meetings twice a year at the university concerned. STW seeks methods to relate the differences in the research outcomes ("evaluation after") to the differences in assessment rankings ("evaluation before"). This study will focus on the evaluation of a sensor technology program managed by STW as a subset of the larger set of all research projects funded by STW.

We go back to the most basic and simple definition of utilization of outcome, namely whether the research results were or were not used by parties outside the university. This simple basis gives surprisingly positive results. First, it does indeed seem that for STW as a whole, the assessment beforehand is a predictor of the chance that the results will be used later. But this does not seem to be true as far as the subset of sensor technology projects is concerned. These findings can help us obtain more insight into what our selection process does and into what determines the success rate in terms of utilization six years after the research has ended.

Overview

The Technology Foundation (STW), a grant-giving agency in the Netherlands, selects research proposals from universities on the basis of their scientific quality and their utilization potential. The proposals relate to research in the field of applied science. STW also assists the research groups in the four years after the grant by bringing together the researchers and the potential users in half-day meetings twice a year at the university concerned. STW seeks methods to relate the differences in the research outcomes ("evaluation after") to the differences in the assessment rankings ("evaluation before"). This study will focus on the evaluation of a sensor technology program managed by STW as a subset of the larger set of all research projects funded by STW.

Within STW, we evaluate research projects "before" and "after." Although we have updated our methods in recent years, we think they can still be improved, especially with regards to the "evaluation after." For this paper we need to define what we mean by "before" and "after." Evaluation "before" means selecting research proposals for funding. STW uses peer review combined with a jury assessment based on a rating according to two criteria: the scientific quality and the utilization potential of a research proposal. The STW Board makes its decisions by giving equal weight to the two criteria, thus ensuring that high-risk proposals are funded. In this study we focus only on the utilization assessment of the jury. Evaluation "after" refers to six years after the research project has ended (ten years after the research started). STW evaluates the outcome of the research in terms of its achieved utilization. So far, STW has used several different evaluation methods. This study attempts to work out a better method.

Last year, STW performed a study (van Caulil et al.1996) to investigate what attempts have been made in other countries to relate evaluations "before" and "after." In that study we tried to arrive at a definition of utilization and to find methods of evaluating completed research projects. We sought to improve our current methods and concluded that there was no ideal method of measuring accurately the relationship between evaluation of the utilization potential before and the actual utilization afterwards. We did not find a set of indicators that exactly measure what we mean by utilization. Although many indicators were found, no good set of indicators for measuring the full spectrum of the utilization outcomes of research was encountered. What we did get is deeper insight into the matter and lots of ideas to work with. On the basis of this knowledge we decided to go back to the most basic and simple definition of the utilization of outcome, namely the clear use of research results by parties outside the university. This simple basis gives surprisingly positive results. First, it does indeed seem that for STW as a whole, the assessment beforehand is a predictor of the chance that the results will be used later. But this does not seem to be true as far as the subset of sensor technology projects is concerned. These findings can help us obtain more insight into

Journal of Technology Transfer Vol. 22 (2): 53–58.

what our selection process does and what determines the success rate in terms of utilization.

Evaluation by STW "Before"

Since 1981, researchers at Dutch universities have been able to apply to STW for four-year research grants in the field of applied science. The principal investigator is free to choose the subject of research. The proposal must describe the way in which the research will benefit science and must also give an idea of the possible utilization of the expected outcomes and state what actions are planned to encourage utilization. For each proposal, STW asks six peers to give well-founded comments. The principal investigator can react to each comment of the peers and STW draws up a document called "a protocol" which contains all the comments and reactions. Twenty of these protocols for 20 completely different proposals in the field of applied science are assessed by a jury of 12. The jury members give a rating for the scientific quality and a rating for the utilization potential of each proposal. The members of the jury act only for one group of 20 proposals; then a new set of 20 proposals is assessed by a completely new jury. The STW Board awards a grant to the proposals for which the jury has given the highest average score. Usually eight out of 20 proposals receive a grant, which means the success rate is 40 percent.

In this study, we focus only on the utilization aspect. The average jury's rating for the funded proposals range from 2 to 5. Each member of the jury uses a scale that ranges from 1 to 9, representing excellent to poor (van den Beemt et al. 1991).

Evaluation by STW "After"

After a grant has been awarded to a research proposal, STW coordinates so-called "users committees" for each project. Twice a year a meeting is organized at the research lab of the university with the group involved. This meeting brings together the researchers and the potential users; the latter are mostly from industry, but also come from research institutes, academic hospitals, and other interested groups in society. STW stimulates the potential users and starts discussions on this matter. This is done so that after the research has ended, STW is well informed about the utilization of the outcomes. One year after completion of the project, STW compiles a one-page report on the utilization aspect (as distinct from the scientific aspect) and the future potential of each project. Ten years after the start of the research (six years after the project has ended), STW updates the utilization report for all the projects involved. In 1993, STW set up a model consisting of three aspects for determining the degree of utilization so that we would be able to view at a glance how many projects had achieved our goals. These aspects are as follows: "the involvement of the user during the research," "the availability of a transferable product," and "the financial benefits for STW resulting from the research results." Each project was given a score on a four-point scale for each of these three aspects: 0 (poor or low), A, B, and C (excellent or high). See Table 1 for more details. The aspects "involvement of the user" and "transferable product" give us more insight into what has been achieved and how successful the project has been, but these aspects cannot guarantee that the outcomes will be used by others. They cannot be used as indicators for utilization "after." Only the aspect "financial benefits for STW resulting from the research" can be seen as a real indicator of utilization. By financial benefits we do not mean the financial support given by third parties (mainly industry) during the research, but the royalties paid by third parties to STW (or sometimes directly to the university) because they make use of the outcomes of research projects funded by STW. This money is usually paid to STW over a period of years after the research project has ended. This indicator, however, refers to only a part of the utilization aspect. There are many results which are well used in industry, but do not bring in royalties to STW.

In this study, I propose to use another indicator for utilization which incorporates all the ways in which third parties can make use of the outcomes of our projects. This is the YES or NO utilization criterion: are the results used or not? The basic definition of utilization YES and NO does not distinguish between degrees of "utilization." We can only evaluate the results reliably in the long term. Determining a YES or NO answer is far easier than assessing utilization on a scale. Everyone has his or her own interpretation of utilization. Both a small use and a big selling product are utilization.

STW gathers YES data from utilization reports one and six years after a project has ended. The STW prepares for each project a one page report and determines the YES or NO utilization status. Half of the one-page report deals with the utilization process and the other half addresses the reasons behind the success or failure. To gain an appreciation for the scope we use to judge YES and NO, let me share three different examples of "utilization NO." First, a young researcher on a STW project started a one-man firm after receiving his Ph.D. His firm collapsed after some time, and a small part of his idea (a new concept for read-out electronics) was used by Phillips, but not on a specific product. Second, a doctoral student doing his Ph.D. on a STW project stopped and moved to the United States for a job offer. He worked on an original idea that is now a world wide hot topic. The last example of "utilization NO" is STW research that received nice results and a prototype from a firm, but was later blocked by a patent elsewhere.

Going into more detail regarding the utilization process, I will define some extra categories, such as "who was the initiator of the grant application" (industry or university), "what kind of outcome is achieved" (product, method, process), "is there worldwide use or only local use," and "does the key-user operate in the Netherlands or abroad." In this way I can detach the use from the degree of use. All

Table 1. 1993 model for assessing degree of technology utilization

1 Involvement of the (potential) user

0 The project has failed because the results are irrelevant.

A The users are interested in the research. The user provides suggestions for the research.

B The user has made a significant contribution to the project, in terms of money, materials, etc.

C The user has clearly participated in the project. The user has made a very large contribution and (usually) has signed a contract for cooperation.

2 Availability of a transferable product

0 The project has failed at the research stageor the research has ended prematurely.

A There is no concrete product. More research is needed to acquire a useful product.This is still the phase of "basic technology." The main form of output hitherto has been a scientific publication.

B A provisional model has been developed and the results can be used. Verification and refinement are needed before an end product is achieved. The user cannot (yet) use the research product independently.

C A concrete product has been developed. This could be in the form of a computer program, a working prototype, or a process description.

3 Financial benefits to STW resulting from the research results

0 Because the project has failed scientifically or because no users have been found, there have been absolutely no benefits, nor are they to be expected.

A There are no benefits from this project. However, future benefits are not ruled out.

B A part of the knowledge has been sold (e.g., a computer program).

C STW receives constant revenues.

Note: For each of the three aspects, the score can be 0, A, B, or C. "Users" are the companies or institutes interested in the STW-financed research.

the aspects of degree of use help us understand the systems that operate in the knowledge transfer process.

Results

Evaluation "Before"

In all this research, we look at STW as a whole (all fields in technology) versus the STW sensor technology program. The "success rate before," defined as the percentage of proposals that received a grant, for STW proposals in the years 1981 until 1986, was 45 percent. From 1986 until 1992 it was 40 percent. By contrast, the "success rate before" for the subgroup of sensor technology proposals was 65 percent in the years 1981 until 1986. This included 9 rejections, 17 grants, and 6 additional grants funded out of extra government budget that scored just below the grants awarded during the normal procedure. From 1986 until 1992, the "success rate before" was 52 percent (23 rejections and 25 grants). In the competition for grants, the STW sensor proposals had a higher than average score within the STW procedure (on applied science).

Evaluation "After"

Here we use the three evaluation aspects: "the involvement of the user during the research," "the availability of a transferable product," and "the financial benefits for STW resulting from the research results." For each aspect, we combined the four scores into only two categories, because the distinction between 0 and A or between B and C is vague, whereas in most cases the distinction between A and B is clear.

For STW as a whole, we make use of the figures as published by STW in its recent "Utilization Report 1985-1995" which deals with 62 projects for which grants were awarded in 1985. We can only use 44 projects for this study, because we submitted the other 18 for assessment and jury ratings to three other specialized organizations with which we have ongoing cooperative agreements. Their ratings do not correspond exactly to ours. For sensor technology, we make use of the 23 projects granted funding between 1981 and 1986. We found that the sensor projects seem to have more involved users. The involvement aspect is important, but not enough to be an indicator for utilization.

Within the sensor projects we see fewer transferable "products" than in STW as a whole. This does not indicate utilization "after," but only highlights one essential aspect of the outcomes of our projects. The aspect is not enough to measure utilization. A transferable product does not mean that industry will actually use the product.

In looking at financial benefits resulting from research

Table 2. Success rate of research proposals

"Before" Year	STW All Fields Success Rate %	STW Sensors Success Rate %	STW All Fields Applications #	STW Sensors Applications #
1986-1986	45%	65%	892	26
1986-1992	40%	52%	1024	48

Table 3. Utilization aspect for "involvement of user"

Involvement of User "During" the 4-Year Period of the Research	All Fields 44 Projects %	Sensors 23 Projects %
0 Not at all A Interested only	20	9
B Users contribute C Users participate	80	91

Table 4. Utilization aspect for transferable "product"

Availability of Transferable "Product" "After"	All Fields 44 Projects %	Sensors 23 Projects %
0 Nothing available A Basic technology	7	17
B Provisional "product" C Concrete "product"	93	83

"after," we see that nearly 20 percent of the projects of both STW "all fields" and STW "sensors" generate royalties. This more or less indicates the minimum percentage of our projects that have led to actual utilization in society.

It is very difficult to draw conclusions about utilization on the basis of the three aspects. Only the aspect "financial benefits" relates directly to utilization, but it certainly does not cover the entire aspect of utilization.

Evaluation "Before" versus Evaluation "After"

Now we relate the average jury score "before" with the most rudimentary indicator for utilization: YES or NO utilization criterion after the project has ended.

This percentage of utilization YES covers all the aspects of utilization and is roughly 4 times as high as the percentage for the aspect financial benefits (some or considerable financial benefits). Here we see that the aspect "financial benefits" covers only about 25 percent of the utilization aspect. For a few years now, STW has pursued a more stringent knowledge transfer policy and I therefore expect that a higher proportion of projects will generate royalties for STW.

STW's objective selection procedure does not discriminate against high-risk proposals in that it gives the same weight to the utilization potential as to the scientific quality. Earlier studies (van den Beemt et al. May 1995; van den Beemt et al. July 1995) have shown that if a proposal is not original enough, the STW jury will rate it much lower than original proposals. So in our selection procedure, high-risk scientifically original proposals are favoured. This is why we are satisfied with the 30 percent utilization NO.

In our investigation, we found five proposals in the STW "all fields" sample and two in the "sensors" sample which at the moment belong to the category "utilization NO," but might well belong to the "utilization YES" in the future. As an indication, we assume that a quarter of these proposals will fall into the category YES later. This can be seen in Tables 6 and 7.

Evaluation "Before" versus Evaluation "After" (in More Detail)

To obtain more insight into the mechanisms underlying the figures on utilization, we take into account three extra sub-groups of the jury score "before." The jury score is on a scale from 1 (excellent) to 9 (poor). Group 2 includes the jury scores ranging between 2.0 and 3.0, Group 3, the jury scores ranging between 2.9 and 4.0 and Group 4, the

Table 5. Utilization aspect for "financial benefits"

Financial Benefits Resulting for the Research "After"	STW All Fields 44 Projects %	STW Sensors 23 Projects %
0 None A Possibly later	80	83
B Some benefits C Considerable benefits	20	17

Table 6. Project score: Pre-grant and 10 years after

Score of Project	"Before" Average–STW Jury Score 1 to 9 (1=excellent)	10 Years "After" Utilization–STW Board % YES	Correction for Possible Utilization in the Future
STW all fields #44	3.4	70%	73%
STW sensors #23	3.6	70%	72%

jury scores ranging between 3.9 and 5.0. Jury figures above 5.0 (projects not funded) are not included in this utilization study, so we deal only with the three groups mentioned above. We relate these jury scores "before" to the utilization YES percentage (%) "after."

For the STW projects as a whole, we see a nice correlation between the jury score "before" and the utilization score "after." But for the subgroup of sensor technology projects we see a different pattern. Possibly there are large differences between the fields of research with regards to utilization. I have an indication that the initiator plays a role in the utilization. I therefore looked at the initiators: university or small high-tech firm which proposes a collaboration with a university, since we only give grants to universities.

The university is the initiator in 27 percent of the STW "all fields" projects, while in 11 percent the initiator is a small high-tech firm. For the sensor technology group, the percentages are 48 percent and 22 percent respectively. The small high-tech firms never give up and seem to have a very high success rate for utilization "after," but we must be careful with this conclusion because of small numbers.

Conclusion

This is only a preliminary study: I still have to work out the details and intend to go into more detail and find out what makes sensor technology so different from other research areas. For this purpose I will make use of subcategories such as: "what kind of 'product' is realized?", and "is the 'product' used worldwide or not?". I will try to categorize the problems that arise during the six-year period from research outcome to commercial product. I want to develop a better evaluation method to use on larger numbers of projects in order to obtain more significant results.

The tables presented show that it is only on the aspect of "availability of a transferable product" that the STW sensor technology projects get lower scores than the whole set of STW projects in all technological fields. STW sensor technology projects do as well as the entire set of other STW projects, but strangely enough it seems more difficult for the jury to forecast the utilization aspect.

However, we must bear in mind that this study is based on small numbers and that we are dealing with a very basic indicator: "utilization YES." The indicator refers to various degrees of utilization including widespread or local use, direct or indirect use, complete or partial use. The utilization YES/NO criterion therefore seems to be a good starting point for a study of the way in which knowledge is transferred from university to industry. The initiator of the project seems to be an important factor in this transfer process and deserves further study.

References

van Caulil, G. F., C. A. M. Mombers, and F. C. H. D. van den Beemt. "Quantifying the Utilization of Research: The Difficulties and Two Models to Evaluate the Utilization of Research Results." *Scientometrics* 37(3), 1996, pp. 433-444.

van den Beemt, F. C. H. D., and Anthony F. J. van Raan. "What do Bibliometrics and Peer Review of Research Proposals Have in Common?" *Technological Innovation and Global Challenges*. Birmingham: July 1995.

van den Beemt, F. C. H. D., and Anthony F. J. van Raan. "Evaluating Research Proposals." *Nature* 375, May 1995, p. 272.

Table 7. Jury assessment before in relation to the utilization afterwards (*The percentage between brackets is the percentage corrected for possible future use.)

Subgroups on the Basis of Jury Score 2.0-5.0 "Before"	STW All Fields Utilization "After" %YES	STW Sensors Utilization "After" % YES	STW All Fields Number of Projects in a Group #	STW Sensors Number of Projects in a Group #
Group 2 (2.0-3.0)	85%	74%	13	4
Group 3 (2.9-4.0)	74%	40% (45%)*	19	10
Group 4 (3.9-5.0)	50% (60%)*	100%	12	9
Total for all groups	70%	70%	44	23

Table 8. Utilization criterion YES in relation to the main initiator of a proposal (*Because of low numbers, score is not valid.)

	STW All Fields (#44)			STW Sensors (#23)		
"Initiator"	#	Utilization % YES	Average Jury Score	#	Utilization % YES	Average Jury Score
University	12	50	3.3	11	64	3.4
"Branch"	14	79	3.4	4	100*	3.6
Large firm	13	77	3.2	3	33*	4.8
Small firm	5	80	3.6	5	80	4.1

van den Beemt, F. C. H. D., and C. le Pair. "Grading the Grain: Consistent Evaluation of Research Proposals." *Research Evaluation* 1(1), April 1991, pp. 3-10.

Author Biography

Frans C. H. D. van den Beemt has been program director of the Technology Foundation STW in the Netherlands since 1981. He has served as chair of the annual 4S conference, sessions on peer review and research evaluation since 1982, and has given numerous lectures on both topics. He has a M.Sc. degree in physical engineering and taught physics and mathematics at Eckhart College, Eindhoven. From 1977 to 1981, he was system engineer at the Machines & System Group in Holec, Ridderkerk.

Journal of Technology Transfer
Subscription Information

The *Journal of Technology Transfer* is provided as a benefit of membership in the Technology Transfer Society. To become a member, please complete the information below and submit this form, with payment, to:

Technology Transfer Society
435 N. Michigan Ave., Suite 1717
Chicago, IL 60611

Member Information:

Name: (Last)____________________(First) ____________

Title: ______________________________

Address: ______________________________

____

____

City: ______________________________

State: ______________________________

Postal Code: ______________________________

Country: ______________________________

E-mail: ______________________________

Phone: (_____) ________________________

FAX: (_____) ________________________

Membership Type (check one):

- o Individual, US ($85 per year)
- o individual, international ($100 per year)
- o Full-time student** ($30)
  ** Documentation required.

- o For-profit organization ($600 per year)
- o Not-for-profit organization ** ($300 per year)
  ** Documentation required.

Payment Method (in US dollars only):

o Check enclosed $ __________ o Master Card o Visa o American Express

Card # ______________________________ Expiration Date ________________

Signature ____

Information for Authors

The ***Journal of Technology Transfer*** publishes articles that inform and enhance the practice of technology transfer. Particular emphasis is given to the following technology transfer issues and concerns:

- experiences regarding successes or learning from failures
- trends in practice, operations, management, or the transfer environment
- evaluation metrics
- public policy developments
- intellectual property and other legal dimensions of technology transfer
- international technology transfer
- management practices
- transfer mechanisms
- new organizational arrangements or experiments
- technology transfer strategies

Manuscripts. Typed manuscripts submitted for review should be double spaced on 8½x11 white paper using a 12-point font to facilitate reviewing and editing. Including tables, diagrams, and references, the manuscript should not exceed 30 pages.

Three copies of the manuscript are needed for the initial review process. Each manuscript submitted for review should have a cover page with the names of the author(s), affiliation(s), complete mailing address(es), telephone and facsimile numbers, and e-mail address(es), if available. The second page should contain a detailed abstract not to exceed 150 words. The abstract should not only overview the paper but also emphasize how it informs and enhances the practice of technology transfer. The manuscript should begin on the third page; that page should have the title of the manuscript but not the name(s) of the author(s). Footnotes should be used only if absolutely necessary. References should appear at the end of the manuscript.

Initially submitted manuscripts need not conform to a particular manuscript style other than those noted above. Authors will be sent a detailed style sheet if their paper is accepted for publication or invited to be revised and resubmitted. Style sheets are available from the co-editors upon request.

Address all editorial correspondence to either:

Dr. Maria Papadakis, Co-Editor
Journal of Technology Transfer
College of Integrated Science and Technology
James Madison University
Harrisonburg, VA 22807
Phone: (540) 568-2790
FAX: (540) 568-2761
E-mail: j-tech@jmu.edu

Dr. Albert N. Link, Co-Editor
Journal of Technology Transfer
Department of Economics
University of North Carolina at Greensboro
Greensboro, NC 27412
Phone: (910) 334-5146
FAX: (910) 334-4089
E-mail: al_link@uncg.edu

Book reviews. All correspondence regarding book reviews should be addressed to:

Dr. Karen K. Coker
Arkansas Department of Health
Center for Health Statistics
Slot #19
4815 West Markham
Little Rock, AR 72205
E-mail: kcoker@mail.doh.state.ar.us

THE TECHNOLOGY TRANSFER SOCIETY

The Technology Transfer Society was Organized in 1975 as a professional development organization. Its mission is to help our diverse membership find and develop the opportunities, knowledge, methods, and relationships necessary to achieve desired end results in technology transfer. This multinational, multidisciplinary society strives to fulfill the mission through a number of services; for example, networking through chapters and meetings; knowledge transfer through publications and tools, Membership is open to individuals in industry, academia (including students), and government.

For information about the Technology Transfer Society or subscription to the Journal, call (312) 644-0828, extension 213.